A Publication of the American Psychological Association

Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Accuracy of Interpersonal Perception—A General Trait? 1
Victor B. Cline and James M. Richards, Jr.

Motivational Induction and the Behavior Correlates of the Achievement and Affiliation Motives 8
Walter R. Reitman

Some Factors Affecting the Reduction of Overt Hostility 24
Leonard Berkowitz

Fantasy Responses to Subliminal Stimuli 33
Michael J. Goldstein and Richard P. Bastian

The Achievement Motive, Goal Setting, and Probability Preferences 36
John W. Atkinson, Jarvis R. Bastian, Robert W. Earl, and George H. Litwin

The Interest Value and Relevance of Fear Arousing Communications 39
Leonard Berkowitz and Donald R. Cottingham

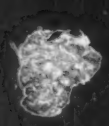
Evaluational Reactions to Spoken Languages 44
W. E. Lambert, R. C. Hodgson, R. C. Gardner, and S. Fillenbaum

Achievement Motive and Test Anxiety Conceived as Motive to Approach Success and Motive to Avoid Failure 52
John W. Atkinson and George H. Litwin

Some Effects of Immediate Versus Remotely Delayed Shock on an Instrumental Response and Cognitive Processes 66
M. E. D'Amato and W. H. Gumenik

Continued on inside front cover

Volume 60 JANUARY 1960 Number 1



CONTENTS—*Continued*

Incidental Stimulation: A Study of Preconscious Transformations 68
Fred Pine

Parental Identification, Acceptance of Authority, and Within-Sex Differences in Cognitive Behavior 76
James Weir

The Effects of Need-Oriented Communications on Attitude Change 80
Francis J. Di Vesta and Jack C. Merwin

Involvement and Communication Discrepancy as Determinants of Opinion Conformity 86
Philip G. Zimbardo

The Effects of Information on the Assessment of Attitudes by Authoritarians and Nonauthoritarians 95
Milton E. Lipetz

Socially Mediated Reduction in Emotional Responses of Young Rhesus Monkeys 100
William A. Mason

Direct and Vicarious Reduction of Hostility 105
Milton E. Rosenbaum and Richard deCharms

Prediction of Small Group Role Behavior from Personality Variables 112
Walter R. Borg

Characteristics and Concomitants of Mood Fluctuation in College Women ... 117
Alden E. Wessman, David F. Ricks, and Mary McIlvaine Tye

Reinforcement of Verbal Affect Responses of Normal Subjects during the Interview 127
Kurt Salzinger and Stephanie Pisoni

Change in Attitudes toward the Negro in a Southern University 131
Robert K. Young, William M. Benson, and Wayne H. Holtzman

CRITIQUE AND NOTES

Some Theory and Data Concerning the Attenuation of Aggression 134
Kenneth R. Wertz

Individual Differences in Reaction to Failure-Induced Stress 136
Samuel C. Fulkerson

A Cross-Cultural Study of Structured Groups with Unpopular Central Members 139
A. J. Mukanna and Michael Argyle

Recalled Dream Themes and Personality 140
Joseph F. Rychlak

Schizophrenic-Like Responses in Normal Subjects under Time Pressure 143
George Usdansky and Loren J. Chapman

EEPS vs Achievement and Verbal Paired-Associates Learning 147
Leonard Worell

Volume 60 January 1960 Number 1

Journal of Abnormal and Social Psychology

Editor: M. Brewster Smith, *University of California, Berkeley*

Associate Editor: E. J. Shoben, Jr., *Teachers College, Columbia University*

Consulting Editors

Gordon W. Allport
Harvard University

Urie Bronfenbrenner
Cornell University

Roger Brown
Massachusetts Institute of Technology

Donald T. Campbell
Northwestern University

Launor F. Carter
System Development Corporation

Richard Christie
Columbia University

Morton Deutsch
Bell Telephone Laboratories

I. E. Farber
State University of Iowa

Donald W. Fiske
University of Chicago

N. L. Gage
University of Illinois

Roger W. Heyns
University of Michigan

Robert R. Holt
New York University

Irving L. Janis
Yale University

Harold E. Jones
University of California

Harold H. Kelley
University of Minnesota

Leonard S. Kogan
Community Service Society of New York

Daniel J. Levinson
Harvard Medical School

David C. McClelland
Harvard University

Leo Postman
University of California

Eliot H. Rodnick
Duke University

Nevitt Sanford
University of California

Seymour Sarason
Yale University

Richard L. Solomon
Harvard University

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 900 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor:

M. BREWSTER SMITH
Department of Psychology
University of California
Berkeley 4, California

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

BARBARA CUMMINGS
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1960 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

ACCURACY OF INTERPERSONAL PERCEPTION —A GENERAL TRAIT?[1]

VICTOR B. CLINE AND JAMES M. RICHARDS, JR.
University of Utah

WHEN in Act I of *Pelleas and Melisande* one of the characters asks, "I have never for one instant seen clearly within myself; how then would you have me judge the deeds of others," a most crucial question is posed for those conducting research in the area of human judgment and interpersonal perception. One of its corollaries would be whether ability to perceive others accurately, if possible at all, is a general trait, or more specific and dependent on *whom* we are perceiving and what sort of judgment of them we are making. In the senior author's prior research in this area (Cline: 1955a; 1955b), the results suggested that ability to judge does have some generality. Later studies by Gage and Cronbach (1955), Crow and Hammond (1957), and Stone and Leavitt (1954), among others, suggested that on the contrary there was little relationship between accuracy of perception scores derived from two or more different instruments or procedures. The present research represents an attempt to determine more precisely if there is a general, global "trait" of perceiving others accurately that is susceptible to meaningful measurement.

METHOD

Subjects

The *Ss* were 50 students, both male and female, in the 1958 summer session at the University of Utah, drawn from classes in psychology, education, sociology, and anthropology on a voluntary basis. Because of the small *N*, results were not analyzed by sex of judges. Such an analysis is planned in later studies. This group was more heterogeneous than is usual in university settings, since it included school teachers and others attending the summer session for special purposes. Although the *N* is rather small for a correlational study, it was thought to be sufficiently large for a study aimed at answering the question of generality.

Stimulus Materials

One of the more difficult problems in studying interpersonal perception is that of holding cues constant for all observers. In our prior research (Cline, 1955a; Cline & Richards, 1958), the most effective and useful procedure for doing this and at the same time providing the observers with a vivid impression of the "stimulus person" was found to be the use of sound motion pictures of an interview situation. Black and white film, employed in the initial stages of this research, proved to obscure too many cues relating to dress, complexion, etc. Therefore all interviews used in the present study were filmed in color.

These interviews as nearly as possible represented a cross section of adults in the Salt Lake City metropolitan area. The interviews were filmed at three different locations: an anteroom to a suburban supermarket, a downtown theater lobby, and a university research laboratory. Interviewees[2] were recruited from supermarket customers, off the street, and from university classes. All interviewees were given a $5.00 merchandise order for participating. All interviewees entered the situation "cold." The camera and sound equipment were in full view of the interviewee during the filming of the interview. Experience with this technique has indicated that their presence does not significantly affect the spontaneity of the interview.

All interviews were conducted by an actor, a member of the University of Utah theater staff. The interview procedure, while permitting the interviewer some freedom, was sufficiently structured to insure equivalence over interviews. The interview probed the following areas: (*a*) personal values, (*b*) personality strengths and weaknesses, (*c*) reaction to the interview, (*d*) hobbies and activities, (*e*) self-conception, and (*f*) temper. Each interviewee was given a brief explanation of the purpose of the research, followed by a "warm-up" period to gain rapport. The interviewer then gave a signal to the camera and sound man to begin filming, and started the interview proper.

Eventually 25 interviews were filmed in sound and color. From this larger group, 10 films were selected for the present research on the basis of technical excellence of sound and color, significant content of the interview, and a pilot study reported in greater detail elsewhere (Cline & Richards, 1958) which employed 20 films. Selection was made on the basis of the extent to which films discriminated good judges from poor judges, and thus closely approximates standard item analysis. Selection on this basis does, of course, to a certain extent "stack the cards" in favor of a general trait. This criticism can be made, however, of any internal consistency item analysis.

The 10 interviewees finally selected included: (*a*) a 19-year-old engaged male geology major at the University of Utah, (*b*) a 36-year-old married female, of Greek

[1] This study was supported by Contract NONR 1288 (04), Group Psychology Branch, Office of Naval Research.

[2] Throughout this paper "interviewee" refers to persons in the films about whom judgments are made.

parentage, with three children, (*c*) a 40-year-old married male occulist of Italian parentage, (*d*) a 22-year-old divorced female waitress with no children, (*e*) a 32-year-old married grammar school teacher with three children, (*f*) a 65-year-old widow with two married children working as supervisor of a church cafeteria, (*g*) a 22-year-old single male Mexican-American working in a meat packing plant, (*h*) a 17-year-old single female high school student of German parentage, (*i*) a 41-year-old married male IBM worker with one child, and (*j*) a 25-year-old divorced female with two children living with her parents. These interviewees are therefore quite diverse with regard to age, sex, and background. No analysis by sex of interviewee was made in the present study, but such analysis is intended in later studies using larger samples.

Judging Instruments

Following the filming of the interview, an intensive study was made of each of 24 interviewees, the 25th being unable to participate further. The following information was obtained in the course of several visits to the interviewee's home. It included an intensive interview on life history and personal habits and the pencil and paper tests listed below:

(*a*) MMPI
(*b*) Strong Vocational Interest Blank
(*c*) California Psychological Inventory
(*d*) a word association test
(*e*) Otis Intelligence Test
(*f*) a multiple choice sentence completion test (taken three times)
(*g*) the Gough Adjective Check List (taken three times)
(*h*) an 80-item cluster from the MMPI (taken three times)
(*i*) a 50-item Likert Trait Rating Scale (self-rating taken three times)

In addition, five close associates were given intensive interviews concerning their knowledge of the interviewee's behavior and history, and they rated the interviewee on the 50-item Trait Rating Scale. These interviews were conducted by a Ph.D. psychologist, a graduate student working toward his Ph.D. in psychology, and a second year law student. All interviews with a given interviewee and his associates were conducted by the same interviewer.

To select these associates, the interviewee was asked to nominate five people who knew him well and for a period of several years. An attempt was made to get both sexes, persons from different generations and status levels (parents, peers, boss, employees, etc.) These five people were then contacted directly by a member of the research team and the interview arranged.

From this information, the researchers constructed five judging instruments for judges to fill out after seeing each filmed interview. These were:

(*a*) ***Behavior Postdiction.*** On this judging instrument, the judge is required to "postdict" real life behavior on the part of the person seen in the filmed interview. A sample item would be:

8. When Mr. M gets angry he usually:
— (*a*) ends up cursing obscenely
— (*b*) gets so upset he has to walk out or leave
— (*c*) tends to keep it to himself and hides his feelings

Items were tailored to each individual separately, and were not common to interviews. The one correct alternative was determined on the basis of the life history interview with the interviewee and his five close associates. Only that behavior where there was agreement by all persons interviewed was used in constructing items. Considerable effort was made to insure that the incorrect alternatives were completely inaccurate and not even partially right. Items were so constructed that in the opinion of the authors a judge could not achieve a high score by always checking the most socially desirable alternative. There was, however, no formal control of social desirability on this or any of the following instruments, since the subjective judgment of the authors was used as the measure of social desirability throughout. Ten items were constructed for each interviewee, making a total of 100. Score on the Behavior Postdiction is the number of correct alternatives checked.

(*b*) ***Trait Rating.*** On this instrument, the judge rated the interviewee on a six-category Likert scale for each of 25 traits. The same 25 traits were used for all interviewees and included both favorable and unfavorable characteristics such as efficient, a leader, impulsive, and unstable. The criterion was obtained from the mean rating on the same trait made by the five close associates of the interviewee. The score on Trait Rating is the average of the squared discrepancies between the judge's ratings on individual traits and the criterion ratings. Cronbach (1955) has developed the rationale for this scoring system extensively. In contrast to the scoring system used for the other judging instruments in this research, this is an error score. In order to prevent confusion, in reporting correlations between Trait Rating and other instruments, the signs have been so arranged that a positive sign means that a good judge as measured by Trait Rating also tended to be a good judge as measured by the other instrument.

(*c*) ***Opinion Prediction.*** This instrument consisted of 20 MMPI items for each interviewee, such as, "I fall in love rather easily." The judge was required to check whether the interviewee answered each question True or False. Only those items that the interviewee answered consistently on each of the three times that the 80-item cluster from the MMPI was administered were included. As nearly as possible these items were selected so that 10 had been answered True and 10 had been answered False. Items were so selected that, in the opinion of the authors, judges could not obtain high accuracy scores on the basis of social desirability alone. Since this scale was constructed from each individual's interviewee's responses, the items included were not necessarily the same for any two different interviewees, although a few items might be common. There was a total of 200 items on this test, and the score is the total number correct.

(*d*) ***Sentence Completion.*** On this instrument, the task for the judge was determining how the interviewee had answered multiple-choice sentence com-

pletion items. Again only those items were included that the interviewee had answered in the same way on each of three separate occasions. On this instrument, as on the preceding ones, in so far as possible items were so chosen that judges could not in the opinion of the authors get a high accuracy score on the basis of social desirability of the answers. There are 100 items on this test, and the score is the total number correct. A sample item is:

4. When I make a mistake, I . . .
   __ (*a*) am embarrassed
   __ (*b*) laugh it off
   __ (*c*) don't give a damn

Since items were selected on the basis of each individual interviewee's responses, the particular items included differed from one interviewee to another.

(*e*) *Adjective Check.* This instrument required the judge to determine which one of a pair of adjectives an interviewee had checked as being descriptive of himself. A sample item is:

18. __ (*a*) Resourceful
    __ (*b*) Cheerful

In each pair the correct choice was an adjective that the interviewee had checked as being descriptive of himself on each of three separate occasions, and the incorrect choice was an adjective that the interviewee had failed to check as being descriptive of himself on these three occasions. In the judgment of the researchers, the adjectives in each pair were roughly equivalent as to social desirability, preventing the use of this variable in obtaining a high accuracy score. There were 20 such pairs for each interviewee, a total of 200 items in all, with the particular pairs included differing from one interviewee to another. The score is the total number correct.

These instruments as reported represent final forms. Several of the instruments have been extensively revised and refined on the basis of previous pilot studies. A more detailed report of these revisions is available elsewhere (Cline & Richards, 1958).

Procedure

All judges in this research made their judgments in groups. At the start of each judging session, the judges filled out forms dealing with necessary background information. They were then given a brief period for inspecting the instruments they would have to fill out. At the conclusion of this period, the film of the first interview was shown. The *S*s then filled out the judging instruments with regard to the first interviewee. The film of the second interview was then shown, and the judging instruments filled out with regard to the second interviewee. This procedure was repeated until all 10 films had been shown and all judging instruments completed. All judges saw the films in the same order. Approximately two hours were required for judges to view all 10 films and complete all judging instruments.

RESULTS

In addition to the scores on the various judging instruments a Total Judging score was obtained. This was computed by converting scores on each judging instrument to standard scores ($M = 50$, $\sigma = 10$) to insure equal weighting and adding across tests. In converting to standard scores, Trait Rating was converted from an error score to an accuracy score.

TABLE 1
RELIABILITIES OF JUDGING INSTRUMENTS

Trait Rating total	.72
Behavior Postdiction	.66
Sentence Completion	.67
Opinion Prediction	.67
Adjective Check	.79
Total Judging	.71

A first question relevant to the problem of generality of judging instruments is that of generality within judging instruments, which is to say their reliability. Here the question is one of whether, for a single judging instrument, a judge who judges the first five films well also does well with the second five films. The various reliability coefficients are presented in Table 1. The reliabilities for the Trait Rating and Total Judging were computed from the Spearman-Brown formula based on variances (Gulliksen, 1950) and a split between the first five films and the second five films. All other reliabilities were computed by using the Kuder-Richardson Formula 20.

There are several important considerations in interpreting these reliability coefficients. Guilliksen (1950) points out that the Kuder-Richardson reliability, when computed from Formula 20, is a lower bound to the reliability. In computing the Trait Rating, a simple first half-second half split was used, rather than an odd-even split. In computing the reliability of Total Judging, tests were combined with unit weight, no attempt being made to weight tests in such a way as to maximize the reliability. Therefore the reliabilities in Table 1 are, in general, lower than the maximum reliabilities for the various instruments.

A second question relevant to the problem of generality of judging ability is that of generality across judging instruments. This has to do with whether or not a good judge on one instrument also tends to be a good judge on another instrument. The intercorrelations, means, and standard deviations of the various judging measures are presented in Table 2. In

TABLE 2
INTERCORRELATIONS AMONG JUDGING MEASURES
(N = 50)

Measure	Trait Rating Total	Behavior Postdiction	Sentence Compl.	Opinion Pred.	Adj. Check	Total Judging	M	σ
Trait Rating total	—						2.4273	.8139
Behavior Postdiction	.30	—					54.37	7.62
Sentence Completion	.48	.47	—				46.72	7.38
Opinion Prediction	.52	.50	.47	—			128.90	10.50
Adjective Check	.65	.24	.58	.54	—		156.90	11.76
Total Judging	.63	.44	.63	.65	.66	—	250.64	38.06

r .05 = .27 r .01 = .35

this table signs involving Trait Rating have been so arranged that a positive correlation means that a good judge on Trait Rating also tended to be a good judge on the other judging measure. The mean and standard deviation for Trait Rating are, however, *error* scores.

All correlations in Table 2 between Total Judging and the various individual judging instruments involve the total judging score with the contribution of the particular instrument under consideration eliminated.

Of the correlation coefficients in Table 2, only one would occur by chance with $p > .05$ and only two with $p > .01$. It therefore seems highly improbable that Table 2 represents chance correlations.

Cronbach (1955) has strongly and effectively criticized the use of global measures of accuracy of interpersonal perception. In the present research, the scoring system for Trait Rating was specifically chosen to permit the more analytical study of accuracy scores suggested by Cronbach. (At the present time Cronbach advocates still another procedure for studying interpersonal perception, even going so far as to suggest that accuracy scores of the type used in this research be discarded altogether. The present authors feel, however, that the type of analysis worked out by Cronbach earlier is the more appropriate for their purposes.) When this procedure is followed, the scores on Trait Rating are analyzed into four components. These four components can be further analyzed into variances and correlation terms, but the finer analysis was not attempted here since the four basic components in difference score form appeared more immediately relevant. (The authors intend to make the further analysis in later studies.) These four components are:

1. *Elevation.* This component represents the extent to which the grand mean of an individual judge, taken over all 10 films and all 25 traits, differs from the criterion grand mean. Cronbach is of the opinion that this component has to do with the way in which a judge uses the rating scale, and has little to do with what is usually meant by "accuracy of interpersonal perception."

2. *Differential Elevation.* This component has to do with the judge's ability to predict the deviation from the grand mean of the mean of individual *interviewees* taken over all traits. It therefore measures the variance of an individual judge's ratings. Cronbach states that in some cases this can be meaningful. In the present study, however, it too seems related mainly to the way in which the judges use the rating scale. It should therefore have little to do with accuracy of perception.

3. *Stereotype Accuracy.* This measures the ability of the judges to predict the deviation from the grand mean of the mean of individual traits taken over all interviewees. This should be an important part of accurate interpersonal perception. It measures what Bronfenbrenner, Harding, and Gallwey (1958) have referred to as Sensitivity to the Generalized Other.

4. *Differential Accuracy.* This component measures the ability of judges to predict differences between judges on each trait considered separately. Cronbach states that this component represents an average over traits. It has been referred to by Bronfenbrenner and his associates (1958) as Interpersonal Sensitivity, and should also be important in interpersonal perception. The actual computational procedure used in deriving this component was the subtraction of the other three components from the Trait Rating total score.

The correlations of these components, together with their means and standard deviations, are presented in Table 3. The data clearly confirm Cronbach's recommendation that global accuracy of perception scores be analyzed.

In Table 3 the correlations between the

TABLE 3
CORRELATIONS BETWEEN TRAIT RATING COMPONENTS AND OTHER JUDGING MEASURES
$N = 50$

Measures	Elevation	Differential Elevation	Stereotype Accuracy	Differential Accuracy	Trait Rating Total Score
Elevation	—				
Differential Elevation	.12	—			
Stereotype Accuracy	.04	.01	—		
Differential Accuracy	.29	.29	.16	—	
Trait Rating Total Score	.25	.24	.16	.25	—
Behavior Postdiction	−.05	−.23	.66	.01	.30
Sentence Completion	−.20	.08	.60	.25	.48
Opinion Prediction	.20	.00	.59	.28	.52
Adjective Check	.05	.23	.44	.56	.65
Total Judging	.16	.10	.76	.53	.63
M	.1015	.1162	.7704	1.4392	2.4273
σ	.1154	.0589	.3911	.5759	.8139

r .05 = .27 r .01 = .35

various components and the Trait Rating Total Score involve the total score with the contribution of the particular component under consideration eliminated. In computing the correlations between these component scores and the Total Judging score, however, it proved to be impossible to eliminate the contribution of the particular component under consideration, since the components represent the relative contribution to the variance of the error scores, a peculiarity of the Trait Rating scoring system. No technique was known to derive the relative contribution of these components to the variance of accuracy scores.

In Table 3 the signs of correlations involving Trait Rating and its components have again been arranged so that a positive correlation means what it would ordinarily be taken to mean. The means and standard deviations are, however, of error scores.

No reliabilities are available for these components because the amount of calculation involving in obtaining them is prohibitive. At the present time, however, a program for a high speed computer is being written to obtain these reliabilities.

DISCUSSION

The reliabilities in Table 1 and intercorrelations in Table 2 tend strongly to indicate that a general, global ability to judge others accurately can be meaningfully measured. The data in Table 3 indicate, however, that this global ability involves a complicated "dynamic" process, or, more precisely, a factorially complex one, and helps to clarify what it is that is general in the global measure. Stereotype Accuracy accounts for a large portion of the generality. However, even after this component is eliminated, considerable generality remains, which seems to be related mainly to the Differential Accuracy component. These two components are relatively independent. It thus appears that a good judge may be accurate because he has an accurate stereotype or because he is able to predict specific differences between individuals (or both).

Bronfenbrenner and associates (1958) have recently come to identical conclusions concerning the two independent components in accuracy of interpersonal perception. Since his research used quite different procedures, it and the present study offer striking confirmation of each other.

The results of this study, in other words, indicate that there is a general ability to perceive others accurately. This general ability, however, consists of two (at least) independent parts: Sensitivity to the Generalized Other and Interpersonal Sensitivity in Bronfenbrenner's terminology, or Stereotype Accuracy and Differential Accuracy in Cronbach's terminology.

It may appear paradoxical (or inappropriate) to conclude that at the same time the ability to perceive others accurately is general and that it consists of two independent components. The authors are of the opinion, however, that these seemingly conflicting conclusions mean that the ability to perceive others accurately is factorially complex and that the independent components reflect this complexity. It is therefore still meaningful to talk about a general trait. There are numerous precedents for this in psychology. Perhaps the clearest example is the variable Intelligence, which has been demonstrated to consist of several independent components but is still treated as a meaningful General trait. In investigations of intelligence, even those studies indicating a G factor have typically also shown several independent group factors.

Since the conclusions of this study differ considerably from those of Crow and

Hammond (1957), it is important to discuss methodological differences that might account for these different conclusions. Crow and Hammond used a wider variety of tasks, including some, such as sociometric measures, which seem only remotely related to what would usually be meant by "accuracy of interpersonal perception." Crow and Hammond also used a different scoring procedure, one that is difficult to evaluate. They are of the opinion that it results in a measure of the Differential Accuracy component. The results of this study indicate that Stereotype Accuracy should be included in studies of interpersonal perception, especially in seeking an answer to the question of generality. In those cases in which Crow and Hammond do deal with Stereotype Accuracy, their results also tend to indicate that this component is both reliable and general. Finally, in the present study the measures went through several refinements through procedures closely approximating standard item analysis. This apparently was not done in the Crow and Hammond study. In view of these differences, the authors hold that insofar as Crow and Hammond's study represents a criticism of naive operationism in the study of interpersonal perception, their point is well taken that the *assumption* of a general judging ability over diverse tasks is inappropriate in the absence of supporting empirical data. In view of the present demonstration of considerable generality over heterogeneous items and heterogeneous persons to be judged, however, Crow and Hammond's results cannot be considered to disprove a general ability to perceive people accurately.

One possible criticism of the conclusions of the present research could be derived from the methodological work of Gage, Leavitt, and Stone (1956). These authors hold that any generality of judging ability is likely to be due to real similarity between the persons judged. This view would suggest that selection of films on the basis of their discrimination of good judges from poor, as was done in this research, would result in selection of films in which the interviewees are more similar to one another than they are to the interviewees in the eliminated films. In order to check this possibility, a special key was prepared for the Trait Rating, based on the mean rating on each trait of the criterion scores of the 10 films included in the study. Then an "Index of Similarity" was computed for each of the 10 best films included and for each of the 10 films eliminated on the basis of the earlier pilot study by scoring the *criterion* ratings for each film against this special key. The two groups of films were then compared with regard to this Index. The results of this comparison (t = 1.694, df = 18) fail to support a significant difference between the groups of films in this respect. The small "chance" difference that does occur, however, is in the direction predicted by Gage, Leavitt, and Stone (1956). Generality of judging ability to the extent found in this study cannot, therefore, be explained on the basis of real similarity to one another of the persons judged.

SUMMARY

Recent investigations have suggested that there is little, if any, generality of ability to judge others accurately. This study indicated that on the contrary it is possible to obtain a measure of the ability to judge others accurately that is reliable and general over a variety of judging instruments. The ability to judge others consists of at least two relatively independent components, *Sensitivity to the Generalized Other* and *Interpersonal Sensitivity*. These components are interpreted as reflecting the factorial complexity of the over-all judging scores. The correlations between the different judging instruments cannot be explained on the basis of the real similarity to one another of persons judged.

REFERENCES

BRONFENBRENNER, U., HARDING, J., & GALLWEY, MARY. The measurement of skill in social perception. In D. C. McClelland (Ed.), *Talent and society*. New York: Van Nostrand, 1958.

CLINE, V. B. An analysis of the performance of 100 Air Force officers on the Social Acuity Test. USAF Pers. Train. Res. Cent. tech. Memo., 1955, Proj. No. 7730, No. AF 18(600)-8, Maxwell AFB. (a)

CLINE, V. B. Ability to judge personality assessed with a stress interview and sound-film technique. *J. abnorm. soc. Psychol.*, 1955, **50**, 183–187. (b)

CLINE, V. B., & RICHARDS, J. M., JR. Variables related to accuracy in interpersonal perception. *USN Ann. Rep.*, 1958, ONR 171-146, Dept. Psychol., Univer. of Utah.

CRONBACH, L. J. Processes affecting scores on "under-

standing of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GAGE, N. L., Leavitt, G. S., & STONE, G. C. The intermediary key in the analysis of interpersonal perception. *Psychol. Bull.*, 1956, **53**, 258–266.

GUILLIKSEN, H. *Theory of mental tests.* New York: Wiley, 1950.

STONE, G. C., & LEAVITT, G. S. *Generality of accuracy in perceiving standard persons.* Bur. of Educ. Res., Coll. of Educ., Univer. of Illinois, 1954. (Mimeo.)

(Received January 9, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

MOTIVATIONAL INDUCTION AND THE BEHAVIOR CORRELATES OF THE ACHIEVEMENT AND AFFILIATION MOTIVES[1]

WALTER R. REITMAN[2]
University of Michigan

McCLELLAND, Atkinson, and their collaborators have developed content analytic systems which, applied to modified Thematic Apperception Test (TAT) protocols, yield scores interpreted as indices of motive strength. Measures of the achievement and affiliation motives obtained in this way have been employed in numerous investigations of memory, perception, and performance as functions of motive strength. Occasional inconsistent findings (cf. Reitman, 1957) and failures to replicate previously obtained results (Birney, 1956)[3] however, suggest a need for further analysis and more precise specification of the basic assessment and validation procedures.

Atkinson and Reitman (1956), investigating Atkinson's conception of aroused motivation as a function of latent motive strength and of environmentally determined expectancies that action will lead to attainment of a motive's goal, predicted and found performance at an arithmetic task to be significantly better in a high n Achievement group than in a low group under an Achievement-Orientation condition designed to maximize expectancies of attaining an achievement goal and to minimize all others. In a related investigation, Reitman and Atkinson (1958) found predictions of arithmetic task performance based on motive scores derived from stories written to the first four pictures of an eight-picture measure to be significantly better than those from stories for the last four pictures, a result they considered to have critical implications for experimental and clinical use of the TAT.

The present study derives from these two previous investigations and affords a further opportunity to evaluate the effects of picture cues, picture order, individual vs. group administration of performance measures, and and achievement, neutral or relaxed-affiliation instructions upon correlations between achievement or affiliation motive scores and performance. Three predictions follow from the conception used by Atkinson and Reitman (1956): (*a*) Significant correlations between motive indices and performance should be positive under all conditions. (*b*) Performance should correlate most highly with indices of n Achievement under Achievement-Orientation conditions. (*c*) Performance should correlate most highly with indices of n Affiilation under the Relaxed-Affiliation condition. A fourth prediction is suggested by Reitman and Atkinson's (1958) results. (*d*) For all experimental conditions under which n Achievement or n Affiliation is significantly related to performance, the motive index based on the first three pictures of the present six-picture measure should yield larger correlations than that based on the last three pictures.

METHOD

All *S*s participated in two separate experimental sessions. At the first session, one of two forms of the thematic apperceptive motive measure was administered. At the second session, *S*s worked at three performance tasks under one of five performance conditions. These five conditions, defined by a combination of group or individual administration of the performance measures with either achievement-orienting, neutral, or relaxed-affiliation instructions, were: (*a*) group, Achievement-Oriented—46 *S*s; (*b*) individual, Achievement-Oriented—53 *S*s; (*c*) group, Neutral—46 *S*s; (*d*) individual, Neutral—50 *S*s; (*e*) group, Relaxed-Affiliation—63 *S*s.

Subjects. A total of 258 males from five sections of the elementary psychology course at the University of Michigan participated in the two parts of this experiment during the fall semester, 1955–56, in fulfillment

[1] This report is based on a dissertation submitted to the Department of Psychology, University of Michigan, in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to the National Science Foundation and the Social Science Research Council for fellowship support while conducting the investigation, and to Daniel Katz, John W. Atkinson, David Birch, William Hays, and E. Lowell Kelly for invaluable aid and advice throughout the course of the Study.

[2] Now at the Graduate School of Industrial Administration, Carnegie Institute of Technology.

[3] Personal communications from R. N. Haber, April 1955 and April 1956.

of the experimental participation requirement of the course.

Measurement of n Achievement and n Affiliation. All *S*s wrote six four-minute thematic apperceptive stories under neutral classroom conditions 10 to 11 weeks before the administration of the performance measures. Two forms were used. The "Ach-pictures first" form consisted of three pictures frequently used to measure the achievement motive followed by three used to measure the affiliation motive. The two three-picture subsets were reversed to make up the second, "Aff-pictures first" form. Picture order within each subset was constant. The three achievement-related pictures used are those designated 1B, 1H, and 1D in McClelland, Atkinson, Clark, and Lowell (1953, p. 375). The affiliation-related pictures used are designated 3E, 3A, and 3B by Atkinson, Heyns, and Veroff (1954, p. 405). Pictures 1B, 1H, 3A, and 3B also were used in the two antecedent studies by Atkinson and Reitman (1956) and Reitman and Atkinson (1958). The first form was administered in three lecture sections (C, S, T), the second in two (A, F). Stories to all six pictures were scored for both the achievement and affiliation motives by experienced scorers, all of whom had obtained scoring correlations above .90 with one of the authors (Atkinson) of the scoring systems in checks of their scoring reliability.

Assessment of performance. To control for interaction between motivational induction instructions and aptitude, Quantitative (Q) and Linguistic (L) Aptitude scores from the ACE examination were obtained for 195 of the 258 *S*s taking part in the experiment; the remaining 63 *S*s were assigned to the Relaxed-Affiliation condition.

Each *S* used a booklet containing one of the three induction instructions and all three performance measures. Each induction instruction consisted of a longer introductory part, forming the first page of the booklet, and briefer "explanatory remarks" about each performance measure. The latter appeared together with the instruction for doing the task on a single page, with the intent of maintaining the motivational orientation induced by the introductory instruction.

The achievement-orienting instruction, stressing the sensitivity of the performance "tests" to care and effort and emphasizing that performance would be evaluated to see "how well you can really do," was designed to arouse a strong expectancy that excellent performance would lead to attainment of an achievement goal, and to minimize all other expectancies. The neutral instruction described the tasks as part of a routine demonstration intended to familiarize students with some typical psychological measures. The relaxed-affiliation instruction, in effect denying the possibility of evaluation by describing the tasks as "preliminary" and "unstandardized," and emphasizing instead *E*'s appreciation of *S*'s helping him "work the kinks out," was designed to arouse an expectancy that performance would be related to attainment of an affiliation goal and to minimize all other expectancies.

The first of the three performance measures was a set of three-step arithmetic problems used in the two antecedent studies. In each problem, two sets of three numbers each are added and subtracted separately; the final solution involves addition or subtraction of the two intermediate results. A spatial perception measure, taken from Thurstone and Thurstone (1943), was next. In each group of items, *S* must discriminate test figures which are rotations of an inspection figure from those which are not. A third measure consisted of items from a Number Matching Test developed by Sperber (1955), in which *S* inspects both members of a pair of numbers and indicates whether they are identical.

Immediately before the first performance measure, *S*s read a short story and were told that they would have to recall it later. This was done in connection with another experiment, and reproduces an experimental procedure common to the two antecedent studies. Two other brief measures and a subsequent questionnaire introduced for other purposes are not analyzed here, except for one questionnaire item checking a hypothesis offered by Atkinson and Reitman (1956, p. 363) about the extent to which recall of the short story may have confounded motive-performance correlations anticipated in their study.

In the individual conditions, *S*s were brought into one of eight small laboratory rooms on either side of a central corridor, one *S* per room. The rooms were closed, *E* conducting the experiment from the corridor. In both individual and group conditions, when all *S*s had arrived, *E* instructed them to turn over the booklets which lay face down before them, read the instructions, and fill in any information requested. To minimize variance due to interaction between *E* and *S*s, *E* did not otherwise communicate with *S*s except to direct them to a new instruction or measure. The *S*s worked nine, four, and three minutes on the arithmetic, spatial, and matching problems respectively. Except for the three Relaxed-Affiliation groups of approximately 21 *S*s each, the average N per session was about seven.

RESULTS AND DISCUSSION

Performance as a function of achievement-orienting vs. neutral instructions and individual vs. group administration. The effectiveness of the printed achievement-orienting instructions is clear from a comparison of performance under the two achievement-oriented conditions with performance under the two neutral conditions. Means and *SD*s of correct performance scores for all five conditions are given in Table 1. Table 2 summarizes the three analyses of variance evaluating performance on each measure as a function of achievement-orienting vs. neutral instructions and individual vs. group administration of the response measures. The printed achievement-orienting as compared with neutral instructions yield a significant increment in number of correct solutions on all three measures ($P < .05$).

TABLE 1
MEAN CORRECT ON RESPONSE MEASURES AS A FUNCTION OF INSTRUCTION AND ADMINISTRATION

Performance Condition		N	Response Measure					
			Arithmetic		Spatial		Matching	
Instr.	Admin.		Mean	*SD*	Mean	*SD*	Mean	*SD*
Ach.	Group	46	42.85	12.46	50.83	22.83	47.35	10.30
Ach.	Indiv.	53	43.87	14.72	56.76	27.51	48.34	11.38
Neutr.	Group	46	36.91	13.04	46.20	19.63	43.70	9.59
Neutr.	Indiv.	50[a]	41.46	13.62	45.70	21.80	45.72	10.44
R.-Af.	Group	63	36.81	14.05	47.52	19.33	45.53	6.90

[a] $N = 48$ for spatial measure; two *S*s in the neutral instruction, individual administration condition did not attempt the spatial perception measure.

TABLE 2
ANALYSES OF VARIANCE OF CORRECT SCORES AS A FUNCTION OF INSTRUCTION AND ADMINISTRATION

Source	*df*	Response Measure					
		Arithmetic		Spatial		Matching	
		MS	*F*	*MS*	*F*	*MS*	*F*
Instr.	1	810.09	4.33*	3223.44	5.82*	475.68	4.24*
Admin.	1	381.90	2.04	363.42	.66	111.93	1.00
I × A	1	151.44	.81	509.49	.92	15.51	.14
Error	191[a]	187.02		553.53		112.17	

*$P = .05$.
[a] $df = 189$ for spatial measure; cf. fn. 5.

Effects of quantitative and linguistic aptitude on performance. As the previous analyses show no evidence of a significant difference between the individual and group conditions, or of interaction between induction instructions and individual vs. group administration, the two administration conditions were combined for analyses of the effects of ACE *Q* and *L* scores. To determine the effects of differences in aptitude, *S*s were divided into three roughly equal groups, first by *Q* and then by *L* scores. Analyses of variance evaluating performance on the three measures as a function of achievement-orienting vs. neutral instructions and of low, moderate, or high *Q* scores show quantitative aptitude to be a highly significant determinant of correct solutions on the arithmetic and spatial measures ($P < .01$) as well as on the matching measure ($P < .05$). Corresponding analyses provide no evidence that linguistic aptitude is related to performance on any of the measures. Finally, neither set of analyses of variance indicates that performance is significantly affected by interaction between induction instructions and either *Q* or *L* scores.

Motive indices and performance. The first prediction states that significant correlations between motive indices and performance should be positive under all experimental conditions. The significance of the motive-performance correlations presented in Table 3 as functions of the several parameters of this study therefore is evaluated with reference to one-tailed significance points.

According to the second prediction, the n Achievement index should correlate maximally with performance under Achievement-Orientation conditions. The tabled correlations do not support this prediction. Of the 12 *r*s between achievement motive scores and the number of items correct on each of the three performance measures under achievement-oriented conditions, only the one with the arithmetic measure for *S*s from Sections A and F working under the individual condition is positive and significant at the $P = .05$ point; 5 of the 12 *r*s are negative.

According to the third prediction, the n Affiliation index should correlate maximally with performance under the Relaxed-Affiliation condition. Of the six correlations between number of items correct on each of the three performance measures and n Affiliation scores based on all six pictures, only one is positive and significant ($P = .05$); four of the six are negative. These results provide no support for the prediction, particularly insofar as a set of larger correlations (up to. 62) between n Affiliation and performance occur under the achievement-oriented, group administration condition, and this is the condition in which such a relationship is least likely to occur according to the theoretical conception that generates the prediction.

The fourth prediction asserts that for all experimental conditions in which n Achievement or n Affiliation is significantly related to performance, the motive index derived from the first three pictures of the present six-picture form should yield larger correlations than that

TABLE 3

CORRELATION OF ACHIEVEMENT AND AFFILIATION MOTIVE SCORES WITH PERFORMANCE AS A FUNCTION OF PICTURE CUES, PICTURE ORDER, MOTIVATIONAL INSTRUCTIONS, AND INDIVIDUAL VS. GROUP ADMINISTRATION OF PERFORMANCE MEASURES

Motive	Motive Score Source and Performance Measure	Achievement-Orientation				Neutral				Rel.-Affil.	
		Aff. Pictures Shown First (Sections A, F)		Ach. Pictures Shown First (Sections C, S, T)		Aff. Pictures Shown First (Sections A, F)		Ach. Pictures Shown First (Sections C, S, T)		Aff. Pic. First	Ach. Pic. First
		Group (N = 21)	Indiv. (N = 23)	Group (N = 23)	Indiv. (N = 30)	Group (N = 24)	Indiv. (N = 20)	Group (N = 22)	Indiv. (N = 30)	Group (N = 31)	Group (N = 32)
Achieve.	Three aff. pic.										
	Arith.	−.18	.26	−.10	−.16	−.18	−.05	.25	−.08	.02	.44[b]
	Spatial	−.02	.36[a]	.04	−.23	−.18	−.12	.16	−.10	−.22	.00
	Matching	−.44	.25	−.20	−.21	−.04	.30	.06	.07	.01	.35[b]
	Three ach. pic.										
	Arith.	.03	.30	.16	−.12	.08	.38[a]	.02	.08	.20	.12
	Spatial	.20	.12	.04	.01	−.08	.14	.08	−.12	−.19	−.17
	Matching	.04	.11	.27	−.12	−.12	.10	.27	.01	.02	−.02
	All six pic.										
	Arith.	−.09	.37[a]	.06	−.18	−.03	.24	.13	.01	.20	.30[a]
	Spatial	.17	.32	.05	−.11	−.15	.04	.13	−.16	−.25	−.13
	Matching	−.24	.24	.09	−.21	−.11	.23	.24	.05	.02	.15
Affil.	Three aff. pic.										
	Arith.	.25	−.27	−.07	.20	.35[a]	−.14	.32	.10	.26	.27
	Spatial	.23	−.14	−.02	.11	−.25	−.23	−.07	.34[a]	.10	−.08
	Matching	.37[a]	−.16	.12	−.07	.14	−.64	.01	−.05	.00	−.12
	Three ach. pic.										
	Arith.	.62[d]	.05	.08	−.15	−.19	−.13	−.05	.01	−.25	.14
	Spatial	.00	−.08	−.29	−.32	−.23	.10	.28	.13	−.18	−.13
	Matching	−.13	.02	−.13	−.01	−.41	−.26	.18	.04	−.12	.14
	All six pic.										
	Arith.	.53[c]	−.22	−.05	.11	.28	−.18	.28	.09	.08	.30[a]
	Spatial	.17	−.15	−.08	−.04	−.29	−.15	.09	.35[a]	−.01	−.11
	Matching	.22	−.14	.09	−.07	.00	−.66	.11	−.03	−.06	−.09
	One-tailed Significance Points for Correlation Coefficients										
	[a] P = .05	.37	.35	.34	.31	.34	.38	.36	.31	.30	.30
	[b] P = .025	.43	.41	.40	.36	.41	.44	.42	.36	.36	.35
	[c] P = .01	.50	.48	.46	.42	.46	.52	.49	.42	.42	.41
	[d] P = .005	.55	.53	.51	.46	.52	.56	.54	.46	.46	.45

based on the last three pictures, regardless of the content of the pictures. Since neither motive index is systematically related to performance under any condition, these data do not permit an adequate test of the prediction. Consider, however, the set of all eight significant correlations between correct scores and motive indices derived from either subset of pictures in either position. The significant correlation is in each case greater than the corresponding one based on the complementary subset of pictures. Three of these eight pairs accord with the prediction, i. e., the larger correlation is based on the subset of pictures presented first; the other five cases contradict it. Finally, as no systematic correlations, either positive or negative, occurred between performance and motive indices, the first prediction also must be considered unsupported.

Further investigation suggests three factors which together may account for the nonreplication in this study of the relationship between n Achievement and performance reported previously (Atkinson & Reitman, 1956).

1. While it is evident from Tables 1 and 2 that for all *S*s combined the level of performance is higher following achievement-orienting instructions than neutral ones, when the com-

parison is made separately by lecture sections, *Ss* from Section C show little difference in performance level between the two instructional conditions. This finding may well be a chance phenomenon, but it also may reflect a real difference in motivational level between these *Ss* and those from other sections.

2. Were TAT administrations comparably "neutral" in the five different lecture sections? Analysis of variance shows that *Ss* from Section S had significantly lower n Achievement scores than *Ss* from the remaining four sections.

3. High n Achievement *Ss* report significantly more effort than low *Ss* to recall the short story during the performance measures under the achievement-oriented group condition. The relatively lesser effort high *Ss* consequently expended on the performance measures may account for the absence of the expected motive-performance relations for *Ss* who worked under this condition.

It is reasonable to speculate that any or all of these three factors may have confounded a relationship between n Achievement and performance in the groups involved, and to emphasize therefore the one remaining set of correlations relevant to the second prediction which are not subject to these three reservations. For *Ss* from Sections A and F who worked under the achievement-oriented, individual condition, the correlation between the six-picture n Achievement index and the arithmetic measure is positive and significant, and the relation of the index to the other two measures is almost as high.

This argument, however, is for two reasons at best only suggestive. It necessarily ignores the over-all chance character of the correlations presented in Table 3, and it depends upon critical variations assumed to exist among *Ss* and/or among measurement conditions. While such assumptions are not infrequent in the literature on n Achievement, they consistently frustrate attempts at rigorous integration of the sizable body of research results now available, and thus constitute a compelling argument for far more interexperimental comparability and standardization of method than now exists in this area.

SUMMARY

Three predictions about correlations between n Achievement or n Affiliation and performance under the conditions of this study are tested. (*a*) Significant correlations between motive indices and performance should be positive under all conditions. (*b*) The n Achievement index should correlate maximally with performance under Achievement-Orientation conditions. (*c*) The n Affiliation index should correlate maximally with performance under the Relaxed-Affiliation condition. A fourth prediction, (*d*), stemming from a related earlier study, specifies that significant motive-performance correlations should be largest when the motive index is derived from the first three pictures of the present six-picture measure. The study also evaluates the effects of several other potential determinants of motive-performance relationships.

One of the two forms of a thematic apperceptive measure was administered to each of five groups of *Ss* under neutral classroom conditions. All *Ss* subsequently participated in a second experimental session 11 weeks later, where they worked at three performance tasks under one of five performance conditions. The Achievement-Orientation conditions yielded a significant performance increment over the Neutral conditions, and this relationship was independent of individual vs. group administration of the response measure, and of differences among *Ss* in quantitative and linguistic aptitude. None of the four predictions were supported by the motive-performance correlations obtained. While a supplementary analysis suggests some slight evidence in support of the second prediction, the results and discussion on the whole serve primarily to demonstrate the urgent need for standardized and more precisely controlled techniques in this area.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

BIRNEY, R. C. Experimenter effect on the achievement motive. Unpublished paper, Amherst College, 1956.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

REITMAN, W. R. Motivational induction and the behavioral correlates of the achievement and affiliation motives. Unpublished doctoral thesis, Univer. of Michigan, 1957.

REITMAN, W. R., & ATKINSON, J. W. Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

SPERBER, Z. A study of the role of anxiety level and defense preference in performance under stress. Unpublished doctoral thesis, Univer. of Michigan, 1955.

THURSTON, L. L., & THURSTONE, T. G. *The Chicago tests of mental abilities.* Chicago: Science Res. Associates, 1943.

(Received September 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOME FACTORS AFFECTING THE REDUCTION OF OVERT HOSTILITY[1]

LEONARD BERKOWITZ
University of Wisconsin

As elaborated by Dollard, Doob, Miller, Mowrer, and Sears (1939), the catharsis hypothesis maintains that the occurrence of an act of aggression by an individual reduces the strength of the aggressive drive within him. In accord with this hypothesis, it has been noted that an initially high level of overt aggression toward frustrating agents frequently is followed by a steady decline in this open hostility (Pepitone & Reichling, 1955; Thibaut, 1950), as if the expression of aggression had "drained" the individual's "reservoir of hostility." However, the catharsis hypothesis is not accepted by all psychologists (e.g., Allport, 1954), and the above mentioned negative correlation between initial and subsequent overt aggression is open to a number of different interpretations. Most obviously, the first acts of hostility could have produced guilt or aggression-anxiety which later caused the person to suppress further outward manifestations of this socially disapproved drive.

Whatever the theoretical interpretation, the decrease in overt hostility following an act of aggression has great practical significance. If such a phenomenon can be produced reliably, for example, legal authorities might be advised to provoke expressions of overt hostility in potentially dangerous mobs in situations in which this hostility can be controlled safely in order to prevent later outbursts when the controlling conditions are inadequate.[2] However, such a technique obviously also has great dangers; the first aggressive acts may provoke a later unpredictable and uncontrollable level of violence (Brinton, 1952).

Some conjectures can be advanced as to the conditions leading to each of these reactions. For example, the first aggressive responses are likely to spark later violent outbursts when (*a*) there is a strong aggressive drive (*b*) that had been inhibited by strong external restraints, and others' reactions to the first expressions of hostility reveal not only (*c*) that the restraints are now considerably weakened, but (*d*) that the aggressive behavior is approved by an appropriate reference group. The present study is more directly concerned with factors affecting the reduction of overt hostility. In this case, it can be proposed that the negative correlation between the intensity of early and later overt aggression is likely to arise when (*a*) the individual reacts to his first aggressive acts with guilt or aggression-anxiety,[3] and/or (*b*) the individual is not further angered either by himself or by others. Alternative (*a*) maintains that the guilt or anxiety produced by the initial display of anger can cause the person to inhibit further hostile acts. According to some suggestive findings (Berkowitz, 1958a), middle class college students frequently respond t their own aggressive behavior with aggression-anxiety even when the overt aggression is socially sanctioned. Thus, this proposition does not necessarily assume the validity of the catharsis hypothesis. The first exhibition of aggression may lead to a subsequent decrease in overt hostility, if the aggressive drive is not further heightened or maintained by additional stimulation, because aggression-anxiety inhibits the later manifestations of hostility.

The present experiment is devoted primarily to a test of these conceptions. More specifically, it deals with the effects of events following the individual's initial display of aggression upon his subsequent level of overt hostility. Obviously, further actions by the perceived instigator (*P*) can either weaken or strengthen the individual's hostility producing frustrations, resulting in either less intense or heightened overt aggression. Moreover, the effects of these later actions by *P* can be augmented by self-stimulations. For example, the individual (*S*) may not "get over" his

[1] This research was supported by research grant M1540 from the National Institute of Mental Health, United States Public Health Service. Thanks are due to Isabel Birnbaum for her invaluable assistance in collecting the present data.

[2] Ethical considerations aside.

[3] This proposition is not necessarily inconsistent with that advanced by Sears and his colleagues (1953) who suggest that frequent punishment for hostile acts leads to aggression-anxiety which, presumably because this anxiety is frustrating, leads to a heightened aggressive *drive*. The present formulation refers to *overt behavior*.

anger quickly, and could even develop a stronger aggressive drive, if he reminds himself of all the frustrations he has suffered at *P*'s hand. Recalling these frustrations, he stimulates himself to maintain or increase his anger toward *P*. If *S* receives some rewards from *P* after expressing aggression towards him, however, this process of self-stimulation could well increase *S*'s aggression-anxiety. Thinking of *P*'s now "decent" behavior, *S* might feel sorry that he had directed hostility towards him.

Two hypotheses follow from this reasoning: (A) Given that *S* had been angered by *P*, the magnitude of his initial increase in hostility has no relationship to any changes in *S*'s hostility level by the conclusion of the episode (or experiment) if *S* is further frustrated by *P* and is reminded of these frustrations. (B) In contrast, the greater the increase in *S*'s overt hostility toward *P* after the initial instigation the friendlier he should become if the first act of aggression against *P* is followed by *P* rewarding *S* in some manner and relatively frequent recall of this reward.

In addition, the present study seeks to test Feshbach's "symbolic catharsis" hypothesis (Feshbach, 1955). According to this view, symbolic aggression expressed in fantasy activities, such as taking the TAT, results in a cathartic reduction of the aggressive drive. But, as has been pointed out elsewhere (Berkowitz, 1958), there are difficulties confronting this hypothesis. An alternative possibility is that the hostility shown by Feshbach's *S*s in response to the TAT cards gave rise to guilt or aggression-anxiety which produced suppression or repression of further manifestations of this drive. There is little doubt that the arousal of a socially disapproved drive can lead to a decrease in the number of overt drive related responses even in fantasy tasks (Chasdi & Lawrence, 1955; Clark, 1955). Thus, there are two additional questions in the present investigation. Can fantasy activities mediate a reduction in overt hostility, and, if so, is this reduction consistent with the hypothesis that guilt inhibits hostility?

METHOD

Procedure and Experimental Manipulations

Two *S*s, both male or female, were brought together ostensibly for a study of "first impressions." The two *S*s were not given an opportunity to talk to each other at this time. After receiving the instructions and a brief description of their task, they went into separate cubicles where each indicated his impressions of his partner on the first questionnaire (Q1). Each was informed that he would have the opportunity to make other evaluations of him after receiving more information about him later. As a working assumption, it is proposed that expression of attitudes of unfriendliness and rejection on the present questionnaires can be regarded as a form of overt hostility.

When the first questionnaire was completed, the *S*s were told to write two notes to each other, at five-minute intervals, describing themselves. They were to utilize their partner's (*P*'s) self descriptions in evaluating him on the second questionnaire. For the first experimental manipulation in the present $2 \times 2 \times 2$ factorial design, the arousal of Initial Unfriendliness, *E* substituted previously constructed notes for the notes actually written. The notes received by the *S*s in the Initial Unfriendliness condition indicated that *P* had little regard for the University and most of its students, while the *S*s in the Initial Friendliness condition were led to believe that *P* liked life at the University, and was fairly active in student affairs. Following this, *S*s again indicated their impressions of *P* on a second form of the questionnaire (Q2).

The remaining manipulations were undertaken to investigate the effects of further stimulation to aggression after the initial arousal. After *E* had collected Q2 the *S*s were instructed that they would now be given still more information about their partners: they would be told what their partners thought of them. *E* gave each *S* a previously completed Q2, saying that this was the partner's second rating of the *S*. In half of the conditions (Unfavorable Evaluation by *P*) the ratings were uniformly low, averaging 3 on the 10-point scale, while the ratings were uniformly high, averaging 8 on this scale, in the remaining conditions (Favorable Evaluation by *P*). Finally, a set of conditions was established in which the *S*s were to stimulate themselves to further aggression toward *P*. After seeing their coworkers' evaluations of them, half of the *S*s were required to fill out a 145-item adjective check list on which they were to indicate the items that characterized themselves. It was assumed that in completing this task these *S*s would continually remind themselves that they did not like *P*, if they had been "aroused" by him, because *P* did not care for *S* or the University.[4] On the other hand, the *S*s who had received favorable evaluations from *P* supposedly would stimulate themselves to an increased liking for *P*. The remaining half of the *S*s were given a non-self-oriented task taking approximately the same amount of time; they were to respond to four TAT cards. These *S*s, it was assumed, would not stimulate themselves, either to like or dislike *P*, to the same extent as the *S*s given the self-description task. (On the other hand, if Feshbach's (1955) hypothesis of a "symbolic catharsis" is correct, the TAT may show less unfriendliness after anger arousal than the *S*s completing the check list.) Following this task, the *S*s again rated *P* (Q3).

[4] The *S*s described themselves instead of *P* for two reasons: (*a*) to remind themselves that it was their self-concept that had been attacked by *P*, and (*b*) unfavorable descriptions of *P* might produce a "cathartic discharge" of hostility.

Dependent Variables and Subjects

All three questionnaires completed by the *S*s contained the same 13 10-point rating scales, but the order of the items differed on Q1. The responses to six intercorrelated items were summed together to yield the unfriendliness score. One of the items reads, "Would you admit him (her) into your circle of close friends?", with the answers ranging from "Definitely yes" (score = 1) to "Definitely no" (score = 10). To summarize, the various ratings of *P* were made at the following times:

After "first impression" (no communication), Q1
After exchanging notes with *P*, Q2
After seeing *P*'s evaluation of them and then responding either to the adjective checklist or the TAT cards, Q3

A second hostility measure was obtained for those *S*s who had been given the four TAT cards. This was a rating on a seven-point scale of the aggression expressed to each card, summed over all cards, using the category system employed by Hokanson and Gordon (1958). The rater did not know the conditions in which the *S*s had been placed. As is the case with the questionnaire scores, the higher the score the greater the hostility.

The *S*s were volunteers from introductory psychology classes at the University of Wisconsin. There are seven women and three men in each condition. The nature of the deceptions and the reasons for them was explained to all *S*s at the completion of the last questionnaire. Most of the *S*s expressed a great deal of interest in the experiment and all readily promised not to talk about the study to other students.

RESULTS

Changes in Unfriendliness

There were no significant differences among conditions, by analysis of variance, in the unfriendliness scores on QI. The means ranged from 19.4 to 24.2. Table 1 presents the mean change in unfriendliness in each of the conditions after the initial arousal (change from QI to Q2), and again after the further stimulation treatments (change from Q2 to Q3). Analyses of variance of these change scores indicate that the initial arousal manipulation was highly successful—there was a considerable increase in unfriendliness from Q1 to Q2 in the Initial Unfriendliness conditions, while the remaining *S*s became somewhat friendlier ($F = 46.46$; $p < .001$). (This was the only significant effect in the variance analysis of the Q1 to Q2 change scores). The *S*s receiving unfavorable evaluations of themselves from *P* also became significantly more hostile from Q2 to Q3 than the *S*s receiving the favorable evaluations. The main effect for Evaluations from *P* is highly significant ($F = 50.89$; $p < .001$), demonstrating the success of the stimulation-from-another treatment.

There also are indications providing at least partial support for the self-stimulation manipulation. First, for the two treatments receiving favorable evaluations from *P*, those *S*s completing the check list, and therefore, who presumably reminded themselves of *P*'s rewards, exhibited a somewhat greater increase in friendliness from Q2 to Q3 than the TAT *S*s. These differences might have become significant if a stronger level of hostility had been aroused in the initial communication period. As will be shown later, the stronger this anger arousal in the check list *S*s (but not in the TAT group), the reliably greater their subsequent change toward friendliness. It may be that the self-stimulation task in this condition caused the *S*s to feel guilty about their previous rise in hostility toward *P*.

In contrast, when *P* had attacked *S* both indirectly through the notes and directly through the unfavorable evaluation (i.e., in the Initial Unfriendliness-Unfavorable Evaluation from *P* condition), completing the adjective check list served to heighten *S*'s unfriendliness toward *P*. Thus, the *S*s describing themselves on the check list became significantly unfriendlier from Q2 to Q3 ($p = .05$ by Duncan multiple range test) than those asked to respond to the TAT cards. As can be seen in the first table, and contrary to the "symbolic catharsis" hypothesis, there was virtually no change from Q2 in this latter group.

From another point of view, Table 1 also suggests that *P*'s evaluation of *S* had its greatest effect on *S*'s attitude towards his "partner" when the evaluation was inconsistent with *P*'s earlier notes. Over all *S*s, the Q2-Q3 attitude changes were greatest when *P*'s evaluations were unfriendly but had been preceded by the friendly notes, or were friendly following the more hostile communications.

This effect is seen more clearly in Table 2. For this analysis, the Initial Unfriendliness Favorable Evaluation and Initial Friendliness Unfavorable Evaluation conditions were combined into a group of *S*s receiving evaluations inconsistent with their prior interactions with *P*. The remaining *S*s comprise a group receiving evaluations that generally could be expected from *P*'s notes. Entered into the table is the number of *S*s in each group whose Q2-Q3 attitude change was consistent with *P*'s evaluation of *S* (i.e., greater friendliness when the evaluation was favorable, greater unfriendli-

TABLE 1
MEAN CHANGE IN UNFRIENDLINESS

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Change from Q1 to Q2:				
Adj. C-L *S*s	9.2[a]	9.5	−4.3	−1.2
TAT *S*s	13.1	9.0	−3.2	−3.3
Change from Q2 to Q3:				
Adj. C-L *S*s	12.3c[b]	−10.6b	11.7c	−3.9ab
TAT *S*s	1.0a	−8.2ab	15.4c	0.4a

[a] Positive change scores indicate an increase in unfriendliness while negative scores signify a decrease in this attitude.

[b] Cells containing the same subscript are *not* significantly different from each other, by Duncan multiple range test (Duncan, 1955), at the .05 level. As mentioned in the text, the only reliable difference in the Q1 to Q2 change scores is the main effect for the initial anger arousal.

TABLE 2
NUMBER OF *S*s CHANGING THEIR ATTITUDE TOWARD *P* FOLLOWING *P*'S EVALUATIONS OF THEM, AS A FUNCTION OF THE CONSISTENCY OF THESE EVALUATIONS WITH *P*'S EARLIER COMMUNICATIONS

	Evaluations Consistent with Preceding Notes	Evaluations Inconsistent with Preceding Notes
Number of *S*s exhibiting a change:		
Consistent with *P*'s evaluations of *S*	27	38
Inconsistent with *P*'s evaluations of *S*, or no change	13	2

Note.—$\chi^2 = 9.92$; $p < .01$.

ness when it was unfavorable). As is readily apparent, a significantly greater proportion of the *S*s receiving the inconsistent treatment changed their attitudes in accord with the evaluations.[5] In other words, *S*s were less strongly affected by *P*'s most recent action toward them when this action was consistent with *P*'s earlier behavior. Conceivably, then, the *S*s "shrugged off" the hostile evaluations when these acts followed the unfriendly notes because unfriendly behavior was more or less expected of him.

However, Table 1 also indicates that the presumed self-stimulation process served to increase unfriendliness toward *P* even when *S* had received both unfriendly notes and an unfriendly evaluation from his partner. As noted earlier, the adjective check list *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition exhibited a significantly greater increase in unfriendliness toward *P* from Q2 to Q3 than the TAT *S*s in this condition. Furthermore, the attitude change difference between the favorably and unfavorably evaluated *S*s receiving the hostile notes is statistically significant for the check list *S*s, but not for the TAT group. Thus, it appears that when *S*s are given a non-self-analyzing task (the TAT) after two unfriendly acts from another, they are not particularly affected by the most recent act. They may have adapted to his hostile behavior. The self-stimulation induced by working on a self-description task, on the other hand, does seem to overcome this adaptation effect. What may have happened is that the *S*s, in performing this latter task, "stirred themselves up" to further anger at *P*.

Further information as to the effects of the self-stimulation manipulation is given by the analysis of the self-descriptions themselves. Using only those adjectives whose classification was agreed upon by 7 out of 10 undergraduate judges, it was found that the check list *S*s receiving favorable evaluations from *P* described themselves in a significantly more favorable manner than the check list *S*s given the unfavorable evaluations ($p = .01$ by Mann-Whitney *U* test). Therefore, since the *S*s receiving favorable evaluations were less likely to have their self-enhancement needs frustrated by *P*, it also is likely that they would be more kindly disposed towards him. According to the present assumptions, in completing the self-description task the *S*s then reminded themselves frequently that *P* had either insulted or rewarded them.

Correlations Between Change Scores

The original formulation governing the present study proposed that the increase in unfriendliness from Q1 to Q2 would be negatively correlated with the change from Q2 to Q3, unless *S* stimulates himself to persist or increase in his hostility toward *P*, because the guilt produced by the rise in hostility inhibits further manifestations of aggression. However, *S* may not feel too guilty if the communications received from *P* had not evoked much anger in him. In other words, the strength of the

[5] In this analysis there are no significant differences attributable to whether the change is increased friendliness or increased unfriendliness.

subsequent guilt reaction is a positive function of the degree of anger arousal. It should be possible, therefore, to relate the magnitude of the Q1-Q2 increase in unfriendliness to the Q2-Q3 changes in this attitude within each of the eight experimental conditions. Condition differences in the strength and direction of this relationship can then be traced to situational effects either heightening or weakening the guilt reaction.

The aggression-inhibition should be weaker, for example, if *S* reminds himself, before completing Q3, of *P*'s attacks upon him. In this case, the guilt feelings either do not arise or they are overcome by the persistent instigation to aggression. Therefore, according to Prediction A, there should be no significant negative correlation for the check-list *S*s in the Initial Unfriendliness-Unfavorable Evaluation from *P* condition. Table 3 reveals that this expectation is fulfilled.

The second specific prediction (B) maintains that the self-stimulation process facilitates the reduction of overt hostility in the *S* receiving a favorable evaluation of himself from *P*. Thus, the check list *S* supposedly would remind himself of his unkind behavior in having "downgraded" *P* when *P* had thought so well of him. Although the self-descriptions give no indication of guilt feelings, there is a significantly negative correlation (tau) between the two sets of change scores for the aroused-then-favorably-evaluated *S*s taking the check list. The greater the increase in hostility after the initial arousal the friendlier the *S*s became after seeing that they had been favorably evaluated by *P*, but only when they presumably reminded themselves of *P*'s high regard for them. The *S*s given the TAT cards did not have this relationship.

Table 3 also presents results which appear to be consistent with Feshbach's "symbolic catharsis" hypothesis. There is a significantly negative rank order correlation between the two sets of change scores for the aroused-and-then-unfavorably-evaluated *S*s who were given the TAT cards. Those *S*s who had the greatest increase in hostility after the initial arousal tended to decrease in unfriendliness after the second arousal treatment, apparently, according to the catharsis hypothesis, as a result of engaging in the fantasy activities.

The correlations between the ratings of TAT hostility and two sets of change scores reported in Table 4 provide further information concerning the hostility reduction process. In particular, they indicate approximately when the *S*s expressing an initially strong dislike for *P* changed in their overt feelings toward him. If the correlation between the Q1 to Q2 change scores and the subsequent fantasy aggression is negative, the reduction in overt hostility obviously occurred between these two sets of responses. On the other hand, if this correlation is nonsignificant and if the correlation between the TAT ratings and the Q2 to Q3 changes is negative, the changes in hostility probably came about some time after the fantasy responses.

As is shown in the table, the reduction in overt hostility after arousal did not occur at the same time in all conditions. In the Initial Unfriendliness-Unfavorable Evaluation condition, for example, there is a tendency for the increase in overt hostility from Q1 to Q2 to be *positively* correlated with the ratings of fantasy aggression. Those who had increased the most in unfriendliness from the first to the second questionnaires were still relatively high in this

TABLE 3

CORRELATION (TAU) BETWEEN CHANGE IN UNFRIENDLINESS FROM QUESTIONNAIRE 1 TO 2 AND CHANGE FROM QUESTIONNAIRES 2 TO 3 (N = 10 IN EACH CONDITION)

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Q3 after:				
Adj. C-L	+.09	−.64[c]	+.09	−.41[a]
TAT	−.52[b]	−.09	−.24	+.19

[a] $p < .10$.
[b] $p < .05$.
[c] $p < .01$.

TABLE 4

CORRELATION (TAU) BETWEEN CHANGES IN UNFRIENDLINESS AND RATINGS OF TAT HOSTILITY

P Evaluates *S*:	Initial Unfriendliness		Initial Friendliness	
	Unfavorably	Favorably	Unfavorably	Favorably
Change from:				
Q1 to Q2	+.49*	+.02	−.56**	+.53**
Q2 to Q3	−.54**	−.05	+.29	+.54**

* $p < .10$.
** $p < .05$.

attitude by the time they responded to the TAT cards. However, there apparently was a reaction to this unfriendliness between the TAT and Q3 so that those *S*s having the highest fantasy aggression tended to become friendlier toward *P*, even though he had evaluated them unfavorably. The five high fantasy aggression *S*s in this condition exhibited, on the average, a 3-unit decrease in unfriendliness, while the five low TAT aggression *S*s increased in hostility toward *P*, on the average, 2.8 scalar units.

These results are to be expected on the basis of the "symbolic catharsis" hypothesis. However, this hypothesis alone cannot predict the results obtained in the Initial Friendliness-Unfavorable Evaluation condition. Here the reaction to the initial expression of overt hostility apparently came prior to the completion of the fantasy task. As is shown in Table 4, those *S*s increasing the most in unfriendliness from Q1 to Q2 tended to have the *lowest* TAT aggression. Clearly, then, the advocate of the catharsis hypothesis must make use of two explanatory mechanisms in accounting for these results: catharsis and some mysterious process that regulates the time of its occurrence. The presumed reaction against one's own unfriendliness did not occur in the "non-aroused" *S*s given the TAT, perhaps because the absence of explicit aggressive responses did not set off the guilt reaction.

DISCUSSION

While the present data do not permit an unequivocal interpretation, it is possible to offer a single explanation for all of the above findings; the view advocated here proposes that each of the negative relationships between initial and subsequent unfriendliness resulted from guilt feelings created by the first hostility reaction.

As indicated earlier, this account in terms of guilt reaction serves as an alternative to Feshbach's (1955) "symbolic catharsis" hypothesis. Where the latter view would maintain that the TAT *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition had reduced their hostility in a fantasy catharsis, the opposing contention presented here argues that the aggressive behavior was merely inhibited because of guilt feelings about aggression. Moreover, this interpretation has the advantage of parsimony. Only a single mechanism, guilt is necessary to account for the present results. In contrast, as was indicated earlier, the adherent to the catharsis hypothesis must utilize two processes in explaining the findings, catharsis and some hitherto unsuspected mechanism governing the time at which this "drainage" will occur.

The differences in the time of the lessening of overt hostility shown in Table 4 certainly appear to be consistent with the guilt arousal explanation. The *S*s displaying hostility reduction relatively early, prior to the TAT, had received only one insult, the unfavorable evaluation, and this evaluation was introduced after *S* was set to be friendly toward *P* by the nonhostile notes. Thus, in the context of the friendly notes, *S* presumably felt guilty for his initial aggression fairly early, perhaps as a direct consequence of his hostile response upon receiving the unfavorable evaluations from *P*. The anger aroused by the evaluation could have been followed by the thought that the evaluation might be justified, ("After all, *P* seems to be a nice guy (or girl)," with the magnitude of the aggression-anxiety initiated by this thought apparently being directly proportional to the increase in hostility toward *P* from Q1 to Q2. In contrast, the *S*s who displayed the hostility reduction following the TAT had been aroused by both the notes and unfavorable evaluations. The hostility arousing notes might have provided a context justifying the aggression felt by *S* after receiving *P*'s evaluation of him, thus delaying the guilt reaction until relatively late.

According to the present interpretation, then, Feshbach's results may be attributable to guilt-arousal in the *S*s given the fantasy task. The aggressive responses made to the TAT cards could have initiated guilt feelings (or aggression-anxiety) within the *S*s for the hostility they felt towards the insulting lecturer, perhaps because in responding to the TAT they became very conscious of this hostility. The *S*s given the nonfantasy task did not have these explicit aggression-cues and therefore did not develop the guilt reaction.

If this reasoning is accepted, it apparently becomes unnecessary to propose that the differences between the adjective check list and TAT *S*s in the Initial Unfriendliness-Unfavorable Evaluation condition stem even partly from the former *S*s' presumed self-stimulation to aggression. As has been just

suggested, they may have had a stronger level of overt unfriendliness toward *P* than the TAT *S*s in the same condition, not because they stimulated themselves to aggression, but because the latter *S*s suppressed their overt hostility on Q3.

Nevertheless, there is some presumptive evidence of a self-stimulation process in the *S*s receiving the favorable evaluations from *P*. In the case of the *S*s taking the check list in these conditions, the greater their increase in unfriendliness from Q1 to Q2 the friendlier they became from Q2 to Q3. It may be that the relatively hostile *S*s in these conditions, in responding to the self-description task, reminded themselves of their "unkindness" in having been unfriendly toward *P* when he had thought so well of them. They, therefore, may have stimulated themselves to increased guilt feelings which produced the manifestations of friendliness on Q3. Thus, if this self-stimulation could have occurred in the *S*s receiving the favorable evaluations, it also may have occurred in the *S*s getting the unfavorable evaluations.

In conclusion, a final point may be made concerning the nature of the frustrations giving rise to aggression. A number of writers have suggested (Berkowitz, 1958, p. 270) that an individual is most likely to become angered by a frustration when this frustration involves the nonfulfillment of an expectation. Data reported earlier provide some support for this contention. Assuming that both the unfriendly notes and the unfavorable evaluations from *P* were frustrating to *S*, the second type of frustration followed the earlier frustration for some *S*s but not others. However, instead of the effects of the two frustrations summating, as Dollard, Doob, Miller, Mowrer and Sears (1939) have predicted, the greatest unfriendliness provoked by the unfavorable evaluation over all *S*s came about when this frustration was inconsistent with *P*'s earlier behavior, (i.e., had been preceded by the relatively nonfrustrating friendly notes). Conceivably, the *S*s receiving this inconsistent treatment had expected friendlier behavior from *P*, so that the nonfulfillment of this expectancy gave rise to a relatively strong aggressive reaction. Moreover, this may be a special case of a more general principle. *P*'s evaluations generally produced the greatest over all change in *S*'s attitude towards him, whether this was an increase in friendliness or unfriendliness, when the nature of the evaluations presumably was unexpected.

SUMMARY

Members of pairs of *S*s, ostensibly engaging in a study of "first impressions," indicated their degree of liking for each other at various times during a 50-minute experimental session, right after first seeing each other (Q1) and then twice more after instigations to aggression. Three sets of instigating conditions were created in a 2 × 2 × 2 factorial design. First, half of the *S*s were caused to have an initial dislike for their partner (*P*) by manipulating the written communications supposedly sent by him. The remaining *S*s received "friendlier" messages. The *S*s then completed the second questionnaire (Q2). Following this, half of the *S*s were told that *P* had evaluated them unfavorably on this latter questionnaire, while the other *S*s were led to believe that *P* had evaluated them favorably. But before the *S*s were given the final attitude questionnaire (Q3), half were asked to respond to four TAT cards, with the remaining *S*s assigned to describing themselves on a long adjective check list. It was assumed that the latter *S*s would remind themselves of *P*'s behavior toward them and thus would stimulate themselves to an increased liking or disliking for *P*.

The results indicate the following:

1. Comparisons among the condition means indicates that *P*'s second act, his evaluation of *S*, had its greatest effect in changing *S*'s attitude toward *P* when this act was inconsistent with *P*'s earlier behavior. Thus, if the evaluation was friendly (favorable) but had been preceded by unfriendly communications from *P* there was more likely to be an increase in friendliness toward *P* than if this favorable evaluation had been preceded by the hostile notes. Similarly, the unfriendly (or unfavorable) evaluation was more likely to produce an increase in anger when the earlier communications from *P* had been friendly than when these messages had been hostile in tone. Since the *S*s probably expected evaluations from *P* consistent with his earlier behavior, it is suggested that a frustration is most likely to

give rise to aggression when this frustration involves the nonfulfillment of an expectation. There also is evidence that the presumed self-stimulation induced by working on the self-description task was effective in further increasing unfriendliness toward *P*, as a result of his unfavorable evaluation of *S*, even when this had been preceded by the hostile communications from *P*.

2. The *S*s receiving favorable evaluations of themselves from *P* showed a negative correlation between increase in unfriendliness from Q1 to Q2 and changes in this attitude from Q2 to Q3, regardless of whether they had been initially angered or not, if they took the adjective check list but not if they responded to the TAT. It is hypothesized that the *S*s in these conditions displaying the greatest increase in hostility after their first impressions, in performing the self-description task, reminded themselves of their unkindness in having been unfriendly toward *P* when he had thought so well of them (as indicated by his Q2 ratings). This self-stimulation to strong guilt presumably led to manifestations of friendliness on Q3.

3. The *S*s given the intervening TAT task showed a negative relationship between the increase in unfriendliness from Q1 to Q2 and fantasy aggression if they had been angered only once (by the unfavorable evaluations of themselves). It is suggested that these *S*s felt guilty for their unfriendliness relatively soon, prior to taking the TAT, so that their guilt produced an inhibition of hostility on the TAT. In contrast, the *S*s angered by both the notes and unfavorable evaluation from *P* showed a negative relationship between fantasy aggression and the subsequent attitude changes from Q2 to Q3. Here it is conjectured that these *S*s did not feel guilty as soon as the other *S*s because the first hostile notes from *P* to some extent served to justify the anger provoked by the evaluation, so that the guilt reaction inhibiting overt manifestations of unfriendliness did not occur until just before the final questionnaire was completed.

REFERENCES

ALLPORT, G. W. *The Nature of Prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257–283.

BERKOWITZ, L. Manifest hostility level and hostile behavior. *J. soc. Psychol.*, in press.

BRINTON, C. *The Anatomy of Revolution.* New York: Prentice-Hall, 1952.

CHASDI, ELEANOR H., & LAWRENCE, MARGARET S. Some antecedents of aggression and effects of frustration in doll play. In D. McClelland (Ed.), *Studies In Motivation.* New York: Appleton-Century-Crofts, 1955.

CLARK, R. A. The effect of sexual motivation on phantasy. In D. McClelland (Ed.), *Studies In Motivation.* New York: Appleton-Century-Crofts, 1955.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and Aggression.* New Haven: Yale Univ. Press, 1939.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–45.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.

HOKANSON, J. E. & GORDON, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 327–333.

PEPITONE, A. & REICHLING, G. Group cohesiveness and the expression of hostility. *Hum. Relat.*, 1955, **8**, 327–337.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

(Received October 3, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

FANTASY RESPONSES TO SUBLIMINAL STIMULI

MICHAEL J. GOLDSTEIN AND RICHARD P. BARTHOL[1]
University of California, Los Angeles

RESEARCH on subliminal stimulation has been largely directed toward showing that social, situational, or personality factors influence a subject's (*S*'s) ability to decode stimuli ordinarily considered as below threshold. The task of *S* typically is to decode briefly presented visual or auditory stimuli in a situation in which there is reason to believe that the stimuli will have differential impact on the *S*s. Since the method of study uses impoverished stimuli that are gradually increased in clarity, conventional psychophysical techniques have been employed. The conclusion that social, situational and personality factors do influence perception has been dependent to a large degree on the appropriateness of these traditional psychophysical methods to the problem at hand. The analysis of Eriksen (1958) suggests that the use of these conventional psychophysical methods makes the motivational interpretation suspect. This argument has been underscored by Goldiamond (1958) who, although writing from a different point of view, presents some very convincing evidence that the response chosen as the indicator of perception has more to do with the obtained results than does the motivational character of the stimuli.

This criticism of methods has stimulated the development of another, less controversial approach to the problem of personality variables in perception. The latest approach, derived largely from the work of Klein, Spence, Holt, and Gourevitch (1958), involves a basic change in the experimental task assigned to *S*. Instead of requiring active decoding of a subliminal stimulus, *S*s are instructed to respond to a clearly supraliminal stimulus. While *S*s respond in this manner, *E* periodically introduces a subliminal or more accurately, incidental stimulus, and notes the effect of the stimulus upon the response to the supraliminal stimulus. If the original methods of study, using threshold measurement, study perceptual learning, then Klein's technique appears to shift the emphasis to the study of *incidental* perceptual learning. The results with this technique suggest that it may redirect the emphasis in perception-personality research to more fruitful ground.

The present paper is concerned with the effect of incidental stimulation upon fantasy. Since it is assumed that incidental stimuli affect behavior by registering outside of awareness, it was hypothesized that such stimuli would most likely influence a response such as fantasy which also is believed to arise from stimuli outside of awareness. Specifically, we predicted that the emotional tone of *S*s' fantasy responses would parallel the emotional tone of the incidental stimulation.

STUDY 1

Method

The present report is based upon two consecutive studies which were of identical design except for one major variable described below. *S*s were volunteers from the introductory psychology course at UCLA who were satisfying a course requirement. Thirty-six *S*s were divided randomly into three equal groups which differed in the type of incidental (subliminal) stimulation presented. A *positive* group received four positively toned words, a *negative* group received four negatively toned words, and a *control* group received nonsense symbols which resembled words when projected rapidly. As the supraliminal fantasy stimulus, four slides made from TAT cards were used. Table 1 presents the words received by the three groups and the TAT cards with which they were projected.

The three basic groups were subdivided into a balanced latin square design with two replications, making it possible to evaluate the main experimental effect as well as any interactions between order of presentation and type of TAT card. *S*s were instructed to make up a story about each TAT picture following the instructions developed by McClelland (1953). These instructions provided a 20-sec. delay between

[1] An abridged version of this paper was originally presented by the senior author as part of the symposium "Subliminal Stimulation and the Social Scientist" at the 1958 APA convention. The present studies were conducted on a grant from the Faculty Research Committee of the University of California at Los Angeles. The authors would like to express their appreciation to Allen E. Edwards, Barbara Rosenquest, Samuel Himmelfarb, and Sue Smith for their assistance in various stages of the studies.

TABLE 1
SUPRALIMINAL TAT CARDS AND INCIDENTAL WORDS PRESENTED CONCURRENTLY, BY GROUP

Group	N	TAT Card Number (Supraliminal)			
		1	3GF	7BM	10
Positive	12	Success	Laugh	Happy	Love
Negative	12	Failure	Cry	Angry	Hate
Neutral	12	////	////	////	////

the presentation of the TAT stimulus and the initiation of a response by *S*. A "dummy" TAT slide, upon which no incidental stimulus was projected, started the series. This was followed by the four experimental cards with overlaid incidental stimuli. *S*s told their stories which were picked up by a hidden tape recorder. Following the four critical slides, *S*s were informed about the incidental stimulation, and two tests were made of the recall of the incidental stimuli. The TAT cards were reprojected one at a time and *S*s were instructed to guess the stimulus words (*a*) on a free-recall basis and (*b*) from a four-word forced-choice test. In the first study, one *S* spontaneously reported seeing the words during the experimental procedure and was replaced by another *S*. Almost all *S*s expressed amazement when informed of the presence of the incidental stimulation.

Method of incidental stimulation. *S*s tested individually were seated 8 ft. from a screen. Two adjacent projectors presented the TAT slides and the incidental stimulation. The projectors were set 4 ft. to the rear and slightly to the left of *S* behind a burlap curtain that hid them from view. The TAT slides were presented in a normal fashion without interruption. The duration of the incidental stimulation was controlled by a large (4 ft. diameter) episcotister which was connected to a constant-speed electric motor with a variable speed transmission. Varying the speed of the motor and the size of the slit in the episcotister gave a wide variety of exposure times. Since *S*'s task was not to actively decode the incidental stimuli, the usual methods of threshold determination were not appropriate. Therefore, settings for the incidental stimulation were determined in a pilot study with a different group of *S*s who were actively engaged in responding to a TAT slide. It was found that an exposure of 3.2 msc. did not distract *S* from making up his story. Neither spontaneous verbalizations nor detailed questioning of *S*s revealed that they had any awareness of the presence of the words overlaid on the TAT slides, although the material was visible to *E*s. This setting was then used for both studies reported in this paper. It was not found necessary, contrary to expectation, to vary the setting on an individual basis. The same setting was thus used for both studies. The incidental stimuli were presented repetitively at a 1/sec. rate providing 20 exposures while each TAT slide was on the screen.

TABLE 2
ANALYSIS OF VARIANCE OF EMOTIONAL TONE RATINGS FOR CLEAR STIMULI

Source	Sum of Squares	*df*	*F*	*P*
Total	117.50	143		
Between Subjects				
Treatment	1.52	2	1.132	N.S.
Order	1.09	3	0.540	N.S.
O × T	5.12	6	1.271	N.S.
Error	16.12	24		
Within Subjects				
Trials	0.13	3		N.S.
Cards	36.04	3	21.881	< .001
Treat. × Trials	2.63	6	1.321	N.S.
Ord. × Treat. × Trials	12.01	18	1.214	N.S.
Error	42.84	78		

Results

The TAT stories were collected by a research assistant and neither of the authors had any direct contact with *S*s. All stories were typed from the tapes with no identifying marks. Ratings of the emotional tone of the stories were then made independently by the authors using Eron's (1950) general rating scale for the emotional tone of TAT stories. The ratings of the two authors correlated .79. Using the mean of the two ratings on each *S*, an analysis of variance was performed, the results of which are shown on Table 2.

It can be seen from the table that the presence of the incidentally presented words failed to influence the emotional tone of *S*s' fantasy productions. The only significant *F* present in the data was between TAT cards, supporting Eron's (1950) earlier finding that the different TAT cards are far from ambiguous in emotional tone.

Since it is possible that a message could be received and yet have no effect upon behavior, an analysis was made of the free verbalizations and the forced-choice responses of *S*s to determine if any stimulus registration had taken place. Using Fisher's exact probability method, an analysis was made of the frequency of correct and incorrect guesses of the incidental stimuli under free recall and forced-choice conditions. The results, by individual card, and over all TAT cards, failed to indicate superior guessing in the two experimental groups when

TABLE 3
ANALYSIS OF VARIANCE OF EMOTIONAL TONE RATINGS FOR AMBIGUOUS STIMULI

Source	Sum of Squares	*df*	*F*	*P*
Total	97.79	143		
Between Subjects				
Treatment	7.41	2	4.333	$<.025$
Order	.53	3	0.019	N.S.
O × T	3.71	6	0.723	N.S.
Error	20.53	24		
Within Subjects				
Trials	.77	3	0.003	N.S.
Cards	16.46	3	10.894	$<.001$
Treat. × Trials	3.46	6	1.145	N.S.
Ord. × Treat. × Trials	5.65	18	0.624	N.S.
Error	39.27	78		

compared to the control group. Even when the free-verbalizations and forced-choice responses were classified merely on the basis of whether they were positive or negative in emotional tone, no differences were found.

STUDY II

The large *F* between the TAT cards suggests that the stimulus properties of the cards inhibit variability among *Ss*. It seemed possible that the fantasy stimuli would have to be much more ambiguous if the incidental stimulation were to have any effect. Therefore, a second study was conducted in which the ambiguity of the TAT cards was markedly increased by adjusting the lens on the TAT projector to maximum diffusion of the projected image. Although it is difficult to describe the resulting pictures quantitatively, qualitatively there was an increased fuzziness of the picture, and a general reduction of clarity of the figures and the nature of their activity. However, it was still clear that the pictures dealt with human-like figures. This technique risked the creation of a more homogenous visual field against which the incidental stimuli might have been more visible. However, the subsequent analysis of the free-verbalization and forced-choice responses did not indicate greater awareness of the incidental stimuli when compared with the clear stimulus group.

Aside from the variation in ambiguity the second study was identical in every way to the first. A second sample of 36 *Ss*, drawn from the same population, was used.

Results

The stories in Study II were analyzed in the same way as in the first study. Table 3 below presents a summary of the analysis of variance of the ratings of emotional tone.

It can be seen in Table 3 that the predicted difference between the three treatment groups is statistically significant ($F = 4.333$, 2 and 24 *df*, $p < .025$). An analysis of the individual means reveals that the positive group gave the most positively toned stories, the negative group gave the most negatively toned stories, and the control group lay between the two. These results occurred in spite of the fact that the increase in stimulus ambiguity did not eliminate the variance due to TAT stimulus characteristics. Although the obtained *F* of 10.894 between TAT cards is not as large as in the first study, it is still highly significant ($p < .001$ for 3 and 72 *df*). Again the implication is that structural characteristics of a TAT card place restrictions upon story tone and that incidental stimuli can influence the story tone only within these restrictions and cannot produce any gross modifications of fantasy. For example, when a negatively toned TAT card was presented, the positive group gave the least negatively toned stories, the negative group gave the most negatively toned stories, and the control group fell in between. A negatively toned incidental stimulus could not stimulate a negatively toned story to a positively toned TAT stimulus, merely a less positively toned story than might ordinarily be told.

It was expected that the presence of differences in the treatment groups would be paralleled by an increase in *Ss'* awareness of the nature of the incidental stimuli. The data were to the contrary, however. Fisher's exact probability method applied to the frequency of correct and incorrect guesses in free recall or in the forced-choice test failed to indicate superior guessing of the incidental stimuli in the two experimental groups when contrasted with the control group.

DISCUSSION

The results of the two studies indicate that the presence of incidental (subliminal) stimuli does not produce a uniform effect upon fantasy responses. When the fantasy stimulus is highly structured and present sufficient stimulus clues for its organization, incidental stimuli are of little consequence. However, as the number of cues available for organization of the fantasy stimulus are reduced, incidental stimuli become part of the effective stimulus complex. It is as if an organism, in reacting to its environment, responds to the hierarchy of perceptual cues in some sort of selective order, with the most prominent and familiar being high in the cue hierarchy. However, when the perceptual cues that are high in the hierarchy are not sufficient for the organization of the stimulus situation, cues of lesser intensity are employed. A visual perception task may be thought of as a special case of any message decoding problem. In certain instances, the manifest content of a written message contains sufficient information for reliable, although not necessarily valid, decoding. Other written messages are more ambiguous and may require information about the writer, such as his sex, age, beliefs, group membership, etc. for decoding. Incidental stimuli appear to provide that kind of information when the more prominent stimuli are insufficient for reliable decoding.

The failure to find differences in free-recall and forced-choice recognition in the second study raises some difficult questions. It was assumed, in the design of the present study, that discrimination of the incidental stimuli must precede any influence upon fantasy. Yet *S*s were unable to give evidence for this sort of discrimination following the fantasy task. Two possibilities arise from this finding. First, it is possible that incidental stimuli cause a response while present on the screen but that they have little sustained effect. A study in progress is designed to provide an answer to this question by determining the effects of incidental stimulation upon the later learning of the same material presented clearly. If material presented as incidental stimuli register outside of awareness, then the future discrimination of these same stimuli should be facilitated when they are clearly visible.

The second possible explanation for the failure to find evidence of word discrimination is that actual discrimination of word form did not take place. It has been suggested by Osgood (1957) in his analysis of motivational factors in perception that "the 'hint' provided by the incomplete stimulus display in a tachistoscope may be sufficient to excite the dominant evaluative component of the total mediational process without being sufficient to excite the total reaction pattern." If this analysis is correct, *S*s in the present study were reacting emotionally to the incidental stimuli and were not making a discrimination of a more precise nature. Thus, the same effect would be obtained whether the specific word projected in conjunction with a TAT card was *success* or any other word with a similar positively toned quality (love, happy, delight) were projected. An analysis of free-recall and forced-choice responses of *S*s failed to indicate that this type of gross evaluative discrimination was taking place. However, since these responses were obtained on a delayed basis, their value in a test of this hypothesis may be questioned. Further research is needed using a broader sample of incidental stimuli of varying evaluative significance before the second hypothesis can be adequately tested.

The success of the method used in the present study suggests that it would be fruitful for further experimentation on problems of personality and perception. The question is no longer "Does *S* see or not see a particular stimulus?" It becomes: "What are the effects of stimulation of differing degrees of clarity or intensity upon subsequent behavior?" As so many philosophers of science have indicated, a change in the way of asking a scientific question frequently results in increasing the probability of finding a meaningful answer.

SUMMARY

Two studies were conducted to determine the effect of incidental visual stimulation (words) upon fantasy responses. Both studies used four TAT slides as the supraliminal stimuli. Three experimental groups were formed according to the emotional tone of the words used as incidental stimuli: positive group, negative group and control (neutral)

group. The control group received nonsense symbols that resembled words.

The results of the first study failed to indicate that the incidental stimuli had any effect upon the emotional tone of the fantasy responses. The data did reveal that the different TAT cards were clearly not ambiguous stimuli and tended to elicit stories of a characteristic emotional tone.

In the second study, the ambiguity of the TAT slides was markedly increased. Analysis of the stories under this condition showed that the incidental stimuli altered the emotional tone. The positive group produced the most positively toned stories, the negative group produced the most negatively toned stories, and the control group produced stories of an intermediate emotional tone. The emotional impact of the different TAT cards was still stronger than that of the incidental stimuli. The analysis of *S*s' guesses about the nature of the incidental stimuli failed to indicate discrimination of these stimuli.

The results are interpreted in terms of an information decoding model in which perceptual cues of differing intensities are used by *S* in an ambiguous situation. Where the message is unambiguous, *S*s react only to the prominent cues in the stimulus situation; however, when a message is ambiguous, *S*s are likely to respond to more peripheral or incidental aspects of the stimulus complex.

The question of whether *S*s are discriminating the full meaning of a message or merely reacting to the evaluative (emotional) aspect of a message is also discussed.

REFERENCES

ERIKSEN, C. W. Unconscious processes. In M. R. Jones. (Ed.), *Nebraska symposium on motivation, 1958*. Lincoln: Univer. Nebraska Press, 1958.

ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64** (9, Whole No. 315).

GOLDIAMOND, I. Subliminal perception: I. Indicators of perception. *Psychol. Bull.*, 1958, **55**, 373–411.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNAH. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

MCCLELLAND, D. *The achievement motive*. New York: Appleton-Century-Crofts, 1953. P. 98.

OSGOOD, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1957. Lincoln: Univer. Nebraska Press, 1957.

(Received November 28, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE ACHIEVEMENT MOTIVE, GOAL SETTING, AND PROBABILITY PREFERENCES

JOHN W. ATKINSON,[1] JARVIS R. BASTIAN, ROBERT W. EARL,[2] AND GEORGE H. LITWIN[3]
University of Michigan

THIS investigation deals with the effect of individual differences in strength of achievement motive on risk taking behavior. It focuses upon (*a*) goal setting as inferred from choices individuals make among tasks that differ in difficulty (see Lewin, Dembo, Festinger, & Sears, 1944), and (*b*) the preferences of individuals in an imaginary gambling situation for options that are equated for expected monetary value but differ in the stated probability of winning (see Edwards, 1953).

The senior author (1957, 1958b) has presented a theoretical model that extends some ideas developed earlier by Escalona and Festinger in the resultant valence theory of level of aspiration (see Lewin, et al., 1944). It accounts for the association between achievement motivation and entrepreneurial activity involving risk taking highlighted in the writings of McClelland (1955, 1958). The theoretical model was initially formulated to explain the results of two experiments dealing with different aspects of risk taking behavior. Atkinson (1958a) found that the level of performance of female college students was significantly higher when the probability of winning a small monetary prize by getting a high score was 1/2 than when the expectancy of winning was either very high (3/4) or very low (1/20). McClelland (1958) found that children in kindergarten and second grade who were judged to be highly motivated to achieve on a graphic expression measure of n Achievement preferred an intermediate degree of risk (or difficulty) more frequently in several different competitive games than children presumed to be low in n Achievement. The latter group more often than the high n Achievement group favored playing it very safe or being very speculative.

[1] This research was supported by a Ford Foundation grant to the senior author for studies of Personality Dynamics. Preliminary results were presented in a paper at the 1957 APA meeting in New York.

[2] Now at Claremont Graduate School.

[3] Now at Harvard University.

The principle of motivation evolved out of these and a series of earlier experiments on n Achievement and performance (Atkinson & Reitman, 1956; Atkinson, 1958b) is very similar to principles proposed by Tolman (1955) and Rotter (1954): Motivation = f (Motive $\times$ Expectancy $\times$ Incentive). However, it differs from some similar formulations in its conception of a motive as a relatively stable disposition of the personality.

It is assumed that in competitive achievement situations, two different motives are invariably aroused by cues that elicit expectancies (subjective probabilities) of success and failure—the achievement motive (n Achievement) and the motive to avoid failure. It is further assumed that the incentive value of success, that is, the amount of satisfaction inherent in successful accomplishment of a particular task, is an inverse linear function of the strength of the expectancy of success. Incentive value of success is assumed to equal $1 - P_s$, where P_s is the subjective probability or measure of the strength of the expectancy of success. Hence for a given individual, the multiplicative combination of Motive to achieve success, Expectancy of success (P_s), and Incentive value of success ($1 - P_s$) will always produce a maximum value when P_s is .50, as in tasks of intermediate difficulty.

It is assumed that the (negative) incentive value of failure is greater the easier the task (specifically, that $I_f = -P_s$). That is, the shame and humiliation accompanying failure of an easy task is assumed always to be much greater than that accompanying failure of an admittedly very difficult task. Given this assumption it can be shown that the multiplicative combination of Motive to avoid failure, Expectancy of failure, and Incentive value of failure also reaches a maximum value for a given individual when P_s is .50. However, in this instance the product of the three variables, which represents the intensity of aroused motivation, has a negative sign implying anxiety and avoidant behavior. If both the tendency

to approach success and the tendency to avoid failure are simultaneously aroused, and if the approach-avoidance conflict that ensues is resolved algebraically as supposed, the behavior of an individual when confronted with an intermediate degree of risk will clearly depend upon which of these two motives is stronger (see Atkinson, 1957, 1958b).

It is assumed that persons having high n Achievement scores as measured by thematic apperception (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958b) or by the French Test of Insight (French, 1958), a modified apperception test, are more likely than persons having low scores to be persons in whom the achievement motive is stronger than the motive to avoid failure. Consequently, it is predicted that persons having high n Achievement scores will prefer intermediate risk (or difficulty) to a greater extent than persons having low n Achievement scores.

The prediction of risk taking behavior of particular persons should obviously be improved when the strength of both the motive to achieve success and the motive to avoid failure are assessed and taken into account. Low n Achievement scores can hardly be viewed as a measure of the strength of the motive to avoid failure. Nevertheless, there is considerable presumptive evidence that the motive to avoid failure is likely to be stronger than the achievement motive in many persons having low n Achievement scores. Atkinson (1953) found that persons having low n Achievement scores show a decreasing tendency to recall interrupted tasks that might be construed as past failures when the instructions for performance of the tasks become more achievement related. This effect has often been called repression of failure. McClelland and Liberman (1949) and Moulton, Raphelson, Kristofferson, and Atkinson (1958) have demonstrated the phenomenon commonly called perceptual defense among *S*'s having low n Achievement scores when presented with failure words tachistiscopically. This result is particularly evident immediately after they have taken some kind of achievement test. Brown (1953) has found that persons in the low n Achievement group are more rigid in solving the *Einstellung* water-bottle problems under achievement test conditions. Raphelson (1957) has found that persons having low n Achievement scores *obtained from thematic apperceptive stories during a stressful achievement test* score higher in Test Anxiety (Mandler & Sarason, 1952) and show a greater amplitude of PGR in the stressful achievement situation then persons having high n Achievement scores. Taking all of this behavioral evidence into account, it is not unreasonable to assume that the *unmeasured* fear of failure is probably relatively stronger in relation to strength of achievement motive in more persons having low than high n Achievement scores.[4] Hence there is some basis for predicting a tendency for a group of persons low in n Achievement to avoid intermediate degrees of risk.

METHOD

Subjects

The data were collected in five laboratory sections of a sophomore-junior level psychology course at the University of Michigan during the fall of 1956. The students in the course were the *S*s and the instructors served as the *E*s. In the present analysis only the data for males will be examined, as the n Achievement measure for females has always constituted a separate problem (see McClelland, et al., 1953 pp. 172–181; Atkinson, 1958b). There were 66 male *S*s, not all of whom, however, were available for all of the separate tests of risk taking behavior.

Measure of n Achievement

Several weeks prior to the experimental session the *S*s were given the Test of Insight developed by Elizabeth French (1958). The test consists of 10 short verbal descriptions of behavior that *S*s are asked to analyze. An example cue is, "Bill always lets the 'other fellow' win." Both Forms I and II, as described by French, were used. Half of the *S*s were given each form. The stories were scored for n Achievement according to the procedure described by McClelland, et al. (1953; also in Atkinson, 1958b) by an experienced judge whose reliability had been established as above .90 on earlier samples.[5] The separate distributions of the male n

[4] Three subsequent studies of achievement motivation and risk taking behavior have used an independent measure of what is here conceived as the motive to avoid failure, the Mandler-Sarason Test Anxiety Questionnaire (1952). The results of these studies provide additional support for the inferences made here concerning fear of failure among the *S*s having low n Achievement scores (see Atkinson & Litwin, 1959; Litwin, 1958; Vitz, 1957).

[5] We wish to acknowledge the assistance of Joan Munson, who scored the protocols. Because so many of the interpretations had no achievement imagery, the distinction between doubtful (TI) and unrelated imagery (UI) was ignored, and unrelated imagery was scored 0 instead of −1.

Achievement scores for the two forms of the Test of Insight were broken at the median and those above the median designated as high n Achievement, those below the median as low n Achievement.

Goal Setting in a Modified Shuffleboard Game

In the experimental session each class (*N*s = 15 to 20) was shown a shuffleboard arrangement. Separate arrangements were set up for males and females. A chalk circle one ft. in diameter was marked on the floor. Fifteen lines, one ft. apart, were marked on the floor in chalk. The closest line was 1 ft. away from the target; the farthest line was 15 ft. away. A chalk number was written beside each line. The number corresponded to the distance from the target.

The instructions given to the class were:

> Today we are planning to play a little shuffleboard to see how good you are at it. Each of you will be given five practice shots. You can take your shots from any line you want and you can move around if you care to.

Then each *S* took five shots with other members of the class looking on. The men shot on one side of the room, the women on the other. There was a good deal of informal banter during the test period. No attempt was made to interfere with or to control the lifelike game atmosphere that prevailed. However, one person did keep track of the shots and hits. When *S* finished he wrote his experimental code number on the sheet recording his shots.

Controlled Practice

The *S*s were then sent out of the room and brought back one at a time for controlled practice. Each *S* was given 10 practice shots from Lines 1, 3, 4, 5, 6, 7, 9, 11, 13, and 15 in a predetermined random order. One male and one female *S* and two recorders were present during this regulated practice. As before, male and female *S*s were run on opposite sides of the classroom.

Estimate of Subjective Probability of Success and Final Test

After all *S*s had finished the practice shots they were again brought in individually and told:

> Now comes the big game to see how good you really are. You will have five shots and can stand anywhere you please. Try to get the highest score. Your score will be the sum of the distances of your hits. Before you begin *we would like your estimate of how many of the 125 people in the class you expect to beat.* (These estimates were recorded.) We have marked the chance of hitting at each of four of the lines. These probabilities are based on the practice shots of all class members.

The following probabilities were written on the floor next to the lines as listed: 2 ft.—.60, 4 ft.—.50, 10 ft.—.30, 15 ft.—.02.[6] Since two games, one for males and one for females, were being played simultaneously, the *E* recorded the data for the males and designated one of the *S*s who had finished to record the data for the females.

Betting Preferences

After the shuffleboard game, the *S*s were asked to indicate their preferences among a series of imaginary bets. A questionnaire was passed out containing 15 paired bets and the *S* was asked to select one bet from each pair. The bets differed in probability of winning but were all of a constant expected monetary value. The probabilities of winning that were used were 6/6, 5/6, 4/6, 3/6, 2/6, 1/6. Two different forms of the betting preferences were used, half of the *S*s getting each form. One form had small value bets; the expected value of these bets was 30¢. The other form had large value bets; the expected value of these bets was $300. A sample pair of bets from the small form is: 3/6 to win 60¢ vs. 5/6 to win 36¢. The 15 pairs represented all possible pairs of the six bets. The brief instruction asked *S*s to imagine that they were in a gambling situation with friends shouting encouragement while they rolled a single die. The stated options were interpreted for them in the instructions, e.g., 3/6 meant if numbers 1, 2, or 3 come up on the six-sided die, you win, etc.

Measures of Subjective Probability of Success

No direct measure of expectancy or subjective probability was available for the shuffleboard game. For the final test shots, the probabilities written on the floor and represented as the practice probabilities of the *S* were taken as the measures of expectancy. However, for the free practice shots not even these were available. Geographic distance from the target seemed the most reasonable and obvious measure of expectancy. To predict from the theoretical model, it is necessary to make the following assumptions: (*a*) subjective probability of success at the one ft. line is nearer to 1.00 than .50; (*b*) subjective probability of success at the 15 ft. line is nearer to 0 than .50. These assumptions justify dividing the 15 lines into equal thirds corresponding roughly to high, intermediate, and low subjective probability of success. For the betting preferences, the stated objective probabilities may be considered the best estimate of expectancy of success.

As a result of a misunderstanding on the part of several of the persons recording the data in the shuffleboard game, only partial data were available for some *S*s. The possibility of using only *S*s on whom all data were available was rejected since it would have resulted in very small *N*s for some comparisons. Hence, all available data were used, even though slight variations in the size of samples resulted for the various aspects of the shuffleboard data. For the same reason, the sample for the shuffleboard analysis is somewhat

[6] The numbers did not accurately represent the obtained probabilities. On the basis of pretesting it was anticipated that the obtained probabilities would range from above .80 to below .20. These obtained probabilities were to be written on the floor. However, the *S*s did so poorly in the controlled practice that even the closest lines had probabilities of hitting that were less than .50. The probabilities that were written on the floor were, as a result, a compromise between what had actually occurred (and hence what would be believed by the *S*s) and the range of difficulty that was expected when the experiment was designed.

TABLE 1

PERCENTAGE OF *Ss* HIGH ($N = 27$) AND LOW ($N = 31$) IN n ACHIEVEMENT TAKING SHOTS AT VARIOUS DISTANCES FROM A TARGET IN A SHUFFLEBOARD GAME

Trial	n Achievement	Distance from Target			Intermediate vs. Very Easy and Very Difficult Combined
		Very Easy 1–5 ft.	Intermediate 6–10 ft.	Very Difficult 11–15 ft.	
1	High	4	56	41	$\chi^2 = 8.18^{**}$
	Low	26	19	55	
2	High	4	56	41	$\chi^2 = 9.93^{**}$
	Low	26	16	58	
3	High	4	56	41	$\chi^2 = 5.33^{*}$
	Low	19	26	55	
4	High	7	52	41	$\chi^2 = 1.01$
	Low	13	39	48	
5	High	11	63	26	$\chi^2 = 5.46^{*}$
	Low	13	32	55	

* $p < .01$ in this direction (one-tailed test).
** $p < .0025$ in this direction (one-tailed test).

smaller than that for the betting preferences for which complete data on all *Ss* were available.

RESULTS

Free Practice Trials

The results for the free practice trials in the shuffleboard game are summarized in Table 1. As can be seen in Table 1, the high n Achievement group preferred to shoot from an intermediate distance. More than half of the high n Achievement group's shots were from the intermediate zone on each of the five trials, while in only one trial were more than a third of the low n Achievement group's shots taken from this zone. The chi square tests comparing high and low n Achievement groups in frequency of shots from the intermediate vs. very easy and very difficult zones combined yield statistically significant differences in the predicted direction on four of the five trials. The low n Achievement group clearly preferred the difficult zone to either the easy or intermediate zones. Among the men high in n Achievement who did not select the intermediate zone, the difficult zone was strongly preferred to the easy zone.

In order to summarize the results for the five practice trials in one statistical test, the mean number of shots from the intermediate zone for the two motivation groups was computed. The mean number of shots from the intermediate zone (6–10 ft.) by the high n Achievement group was 2.81 and the mean for the low n Achievement group was 1.32. The difference between these means is statistically significant ($t = 2.92$, $df = 56$, $p < .005$ in the predicted direction). In addition, we can compare these means with the figure we would expect if the distribution of shots across the three zones was rectangular or random.

The chance number of shots expected from each zone on five trials would be 1.67. The mean number of shots taken by the high n Achievement group from the intermediate zone is significantly higher than 1.67 ($t = 2.63$, $df = 26$, $p < .02$), but the difference between the chance figure and the low n Achievement group mean of 1.32 is not statistically significant ($t = 1.25$, $df = 30$, $p < .25$).

Reported Expectancy of Success

There are no significant differences between the number of successful shots for the two motivation groups on either the free practice shots, when distance from the target was held constant, or on the regulated practice shots. Nevertheless the high n Achievement group reported that they expected to beat more persons than the low n Achievement group. A median test shows that 68% of the high n Achievement group ($N = 28$) expected to beat 63 or more persons in the class, while only 35% of the low n Achievement group ($N = 31$) expected to beat this many. The difference is statistically significant ($\chi^2 = 6.17$, $df = 1$, $p < .02$).

Final Test Shots

On the final test shots, the actual probabilities of hitting observed in practice and the probabilities written on the floor were such that if the theory is correct the high n Achievement group would take shots from lines closer to the target than the low n Achievement group. On the basis of the probabilities written on the floor, the region from 2 ft. to 6 ft. would be the .60 to .40 (intermediate) probability region. On the first trial of the test shots, 64% of the high n Achievement group ($N = 25$) shot from 2 ft. to 6 ft. while only 39% of the low n Achievement group ($N = 23$) shot from this region. No one shot from the one ft. line. The difference is statistically significant ($\chi^2 = 2.97$, $df = 1$, $p < .05$ in the predicted direc-

tion). Furthermore, five of the low n Achievement group but none of the high n Achievement group shot from 15 ft. on the first trial.

The two groups did not differ significantly on later trials, probably because the high n Achievement group which hit the target more often while standing closer began to move back. The percentage of high n Achievement *Ss* ($N = 25$) that shot from 2–6 ft. on Trials 1–5 respectively was 64%, 52%, 48%, 44%, 48%. The percentage of low n Achievement *Ss* ($N = 23$) that shot from 2–6 ft. on trials 1–5 respectively was 39%, 39%, 43%, 43%, 43%. Earlier studies (e.g., Atkinson & Reitman, 1956; Mandler & Sarason, 1952) provide sufficient basis for expecting that the high n Achievement group would perform better in a test situation than the presumedly more anxious low n Achievement group. The high n Achievement group hit the target on 30% of its shots; the low n Achievement group hit on only 16% of its shots. The difference between the mean number of hits by each group, using a log transformation to remove skewness, is statistically significant ($t = 2.42$, $df = 46$, $p < .01$ in this direction). When distance of shots is held constant, the high n Achievement group still tends to hit more often than the low n Achievement group. For example, 25.8% of 66 shots by the high n Achievement group vs. 10.7% of 56 shots by the low n Achievement group were hits from 6–10 feet. But the unequal number of shots from various distances by *Ss* in the two groups makes it difficult to clearly evaluate the statistical significance of the difference holding distance constant.

Betting Preferences

In the analysis of the betting preferences, the bet most often selected was considered the preferred bet. There was a tendency on the part of all *Ss* to prefer the 6/6 option, which was money already won. There was no difference between the high and low n Achievement *Ss* in their willingness to accept the sure thing. In the small expected value bets 47% of the high n Achievement group ($N = 17$) and 42% of the low n Achievement group ($N = 12$) clearly preferred the 6/6 option. The chi square test of *Ss* who clearly prefer 6/6 vs. those who showed any other preference does not allow rejection of the null hypothesis ($\chi^2 = .08$, $df = 1$, $p > .75$). For the large expected value options, 33% of the high n Achievement group ($N = 15$) and 41% of the low n Achievement group ($N = 22$) clearly preferred the 6/6 option. Again the null hypothesis cannot be rejected ($\chi^2 = .22$, $df = 1$, $p > .50$).

TABLE 2

PERCENTAGE OF *Ss* HIGH AND LOW IN n ACHIEVEMENT SHOWING A PREFERENCE FOR INTERMEDIATE PROBABILITY BETS (4/6, 3/6, OR 2/6) OR EXTREME PROBABILITY BETS (5/6 OR 1/6) OR NO PREFERENCE IN TWO IMAGINARY DICE GAMES

Expected Value of Options	n Ach	*N*	Probability Preferences			Intermediate vs. Extremes[a]
			Intermediate (4/6, 3/6, or 2/6)	No Pref.	Extremes (5/6 or 1/6)	
$ .30	High	17	53	18	29	$p = .04$
	Low	12	17	17	67	
$300	High	15	33	33	33	$p = .16$
	Low	22	18	18	64	
Both groups combined	High	32	44	25	31	$\chi^2 = 7.43$
	Low	34	18	18	65	$p < .005$

[a] All significance tests are for one-tailed Fisher exact probability tests.

Since the 6/6 option is a riskless option, no prediction is made from the theoretical model about the tendency of persons to prefer it. For this reason, and since there was no difference between the two motivation groups in their tendency to prefer a sure thing, the second most preferred bet was taken as the preferred bet involving *risk* except when some bet involving risk was initially preferred even to the 6/6 option.

If an *S* preferred the 4/6, 3/6, or 2/6 bet to both 5/6 and 1/6, he was classified as preferring intermediate risk. If he preferred the 5/6 or 1/6 bet to 4/6, 3/6, and 2/6, he was classified as preferring extreme risk. If he preferred one of the intermediate bets and one of the extreme bets equally often, he was classified as showing no clearcut preference. Table 2 contains a summary of the preferences of the two motivation groups. For the small expected value bets (30¢), the predicted difference between the two motivation groups in tendency to prefer intermediate vs. extreme risk is statistically significant ($p = .04$ in the predicted direction). For the large expected value bets ($300) the difference between the

two motivation groups in tendency to prefer intermediate vs. extreme risk is in the predicted direction, but is not statistically significant ($p = .16$ in this direction). When the two conditions are combined, the chi square for the difference between the high and low n Achievement groups yields evidence of a substantial difference ($\chi^2 = 7.43, df = 1, p < .005$ in the predicted direction).

When the expected value is small (30¢), the eight *Ss* in the low n Achievement group who prefer extreme probabilities are divided evenly between 5/6 and 1/6. However, when the expected value is large ($300), 10 of the 14 *Ss* in the low n Achievement group who prefer extreme probabilities choose the 5/6 alternative.

A so-called "goodness of fit" test[7] shows that it is the low n Achievement group that shows a strong probability preference; the preferences of the high n Achievement group are not different from what would be expected by chance. Since three of the five bets are intermediate risk bets, .6 of the preferences should be for intermediate risk bets by chance. For the small expected value bets (30¢), the high n Achievement preferences are not different from what is expected by chance ($p = .97$), but the low n Achievement group clearly prefers the extreme risk bets more often than expected by chance ($p = .02$). For the large expected value bets ($300) the high n Achievement group's preferences are again not different from a chance level ($p = .73$), but the low n Achievement group exhibits a strong preference for extreme risk bets ($p = .003$).

Similar Risk Taking Strategies on Two Games

Table 3 contains the percentage of *Ss* high and low in n Achievement who could be classified as intermediate or extreme on both the shuffleboard and betting preferences, or who showed a mixed strategy. To increase the size of the sample, the criterion of intermediacy for betting was that one of the intermediate bets was preferred as often as, or more often than, either of the extreme bets. This is a less stringent criterion of preference that approximates a median break of the preferences. For the shuffleboard game the criterion of intermediacy was that two of the five shots were taken from 6–10 ft.; again, a somewhat relaxed criterion approximating a median break. Table 3 shows that the motivation groups are quite consistent in their approach to the two games. The very significant chi square ($\chi^2 = 10.16, df = 1, p < .001$) shows that of the people exhibiting similar strategies in the two games, those preferring intermediate risk tend strongly to be high in n Achievement and those showing preference for extreme risk tend to be low in n Achievement.

TABLE 3

Percentage of *Ss* High ($N = 26$) and Low ($N = 29$) in n Achievement Showing Similar or Dissimilar Strategies in the Shuffleboard Game and the Betting Preferences

n Achievement	Strategy			Intermediate vs. Extreme
	Intermediate in Both[a]	Mixed	Extreme in Both	
High	46	42	12	$\chi^2 = 10.16$
Low	14	41	45	$p < .001$[b]

[a] Criterion for shuffleboard was 2 of the 5 shots from 6 to 10 ft; criterion for betting preferences was that an intermediate probability bet was preferred more often than or as often as an extreme probability bet.

[b] One-tailed test.

Discussion

The most important feature of this study is the confirmation of the hypothesis derived from the risk taking model. Men who are high in n Achievement show a fairly strong preference for intermediate risk in a game requiring activity in which the outcome is contingent upon successful exercise of skill and a relative preference for intermediate risk in a game of chance where their own skill and competence cannot control the outcome. Men who are low in n Achievement generally prefer extreme probability alternatives in the same games. This is indeed what would be predicted from the model if we knew that in one group the motive to achieve was stronger than the motive to avoid failure, and in the other group the motive to avoid failure was stronger than the motive to achieve. However, in a college sample, the distribution of n Achievement scores probably represents a very selected group of motivated people. Thus it would seem probable that not all *Ss* below the median

[7] Binomial Test (Siegel, 1956) for "goodness of fit." All probabilities are two-tailed probabilities.

in n Achievement have motives to avoid failure that are absolutely stronger than their motives to achieve. It need be assumed only that the avoidance motive is relatively stronger in this group than in the high n Achievement group to predict a relative difference in strength of preference for intermediate vs. extreme risk. For this reason there is little difficulty in making sense of the finding that in the initial "free" shots of the shuffleboard game the men who were high in n Achievement took significantly more than a chance number of shots from 6 to 10 ft., while the low n Achievement group took fewer than a chance number from this region, although the tendency to avoid, i.e., to depart from the chance frequency, was not statistically significant. The more surprising finding is that the low n Achievement group showed any concrete evidence of a tendency to avoid intermediate risk.

As originally designed, the experiment called for a set of final test shots with objective probabilities obtained from practice trials ranging from very high to very low written on the floor. We anticipated even better results with clearly stated probabilities than in the free shots where the probability of success had to be inferred as an initial reaction to distance from the target. But the *S*s were such poor shots, and their inadequacies during practice were so obviously apparent to them, that it was necessary to make the test probabilities believable. Even so, it was necessary to exaggerate their success to get the probabilities written on the floor for the nearest lines over .50. We then made the comparison between groups that the hypothesis called for, and on the first test trial again found it confirmed even though now the high n Achievement group was physically closer to the target (i.e., in the 2–6 ft. range) than the low n Achievement group.

In the shuffleboard game, both the high and low n Achievement groups preferred the very difficult region to the very easy region. This preference can be understood in light of an additional assumption proposed earlier by Atkinson (1957, 1958a) and confirmed by other evidence in the present study: a strong motive either to approach or to avoid some goal may bias the subjective probability of the outcome in the direction consistent with the motive. In other words, the person who is high in n Achievement may have a stronger expectancy of success than is warranted by the objective conditions; and the person in whom the motive to avoid failure is very strong may consistently underestimate his chance of succeeding. The statistically significant difference between high and low n Achievement groups in estimates of the number of persons they *expected to beat* in the final shuffleboard test is clearly consistent with the assumption that there might be motivational biasing of subjective probabilities. The high n Achievement group was more expectant of success than the low n Achievement group even though up to that point in the experiment there was no difference in their level of performance of the task.[8]

If the range of distances in the shuffleboard game represents a symmetrical range of difficulty around a probability of .50, the .50 point would fall roughly in the middle, at about 7 or 8 ft. The low n Achievement group, presumably stronger in the motive to avoid failure, would bias the probability downward and their subjective .50 point would be closer, possibly at 5 or 6 ft. Clearly then, the very difficult region, 11–15 ft., would represent the most successful avoidance of the intermediate probability region. The median of the distribution of all shots by low n Achievement *S*s was at 10.92 ft. The high n Achievement group, on the other hand, should bias the probability upward, and their subjective .50 point should fall farther away, possibly at 9 or 10 ft. The point of maximum preference for the high n Achievement group should then fall in the intermediate zone, but nearer to the very difficult region, 11–15 ft., than to the very

[8] It is worth noting, at this point, the distinction to be made between operations that provide an estimate of subjective probability, e.g., how many do you expect to beat, or how many times out of 10 do you expect to hit the target, and operations that reveal the resultant motivation of the *S*, viz., the level of difficulty actually chosen or the specific goal the individual sets to *try for*. The appropriate use of these often confused operations is illustrated in the paragraphs above. An individual's estimate of the number of persons he expects to beat is used as a rough measure of his subjective probability of success; and the level of difficulty he actually chooses in shooting is used as evidence of his resultant motivation (or level of aspiration).

easy region, 1 to 5 ft. The median of all shots by the high n Achievement group was at 10.03 ft. In other words, the distribution of shots by the two groups shown in Table 2 is consistent with these conjectures regarding possible biasing of subjective probabilities.

The tendency of relatively more anxious *S*s to avoid intermediate risk by preferring "long shots" more frequently than "sure things," here interpreted in terms of biased probabilities of success, may on certain occasions be explained in other ways. The theoretical model (see Atkinson, 1957) clearly states that extreme subjective probabilities (e.g., .10 and .90) which are equally distant from a subjective probability of .50 are equivalent in arousing weak anxiety about failure. However, the model does not necessarily imply that therefore an equal number of *S*s should avoid in each direction. Trying very difficult tasks, even when there appears to be no chance of success, may often appear more socially desirable (Edwards, 1957) than attempting the easiest tasks. Or the number of individuals who avoid failure by excessively high or low levels of aspiration may depend upon the relative frequency of persons in the sample who have developed one or the other of these strategies as a characteristic mechanism of defense. Several writers (Eysenck, 1955; Eysenck & Himmelweit, 1946; Miller, 1951) have shown that neurasthenics and hysterics are characteristically different in just this respect; the former normally set excessively high and the latter excessively low levels of aspiration. Furthermore, we should expect relatively anxious *S*s to pile up at either the very easy or very difficult ends of a risk continuum whenever the range of tasks presented is less than a symmetric range of probabilities around the subjective probability of .50. For example, in the test trials of this experiment, there were no "easy" tasks to choose. The closest lines had stated probabilities of success near .50, so the relatively more anxious *S*s avoided shooting from up close. They appeared to have higher levels of aspiration than the high n Achievement group. The result could have been just the reverse if tasks had ranged from "very easy" (i.e., very high P_s) to "moderately difficult" (i.e., P_s near .50) with no very difficult tasks to choose from.

In other words, we may expect to find n Achievement groups with higher, lower, or intermediate aspirations in relation to those of low n Achievement groups depending completely on the range of difficulty that is presented to the *S*s. It sounds very slippery indeed, to say theoretically anything can happen depending upon the conditions. Yet that is precisely what the theoretical model says.

The obvious problem for future research is to provide satisfactory independent measurement of subjective probability of success. The use of geographic distance to estimate strength of expectancy of success, although the only available measure during the free shots, may be considered the weak point of this investigation. A subsequent study by Litwin (1958) has begun to explore the systematic use of different operations for defining subjective expectancies.

In the betting preferences, the high n Achievement group preferred the intermediate vs. extreme risk bets significantly more often than the low n Achievement group when the expected monetary value was only 30¢. However, when the expected monetary value is increased to $300, the differential in risk taking preferences is reduced. Apparently an increase in the amount of money to be won—even in imagination—increases the relative strength of motivation for money qua money. In the small expected value bets, the amount of money may function as no more than a symbol of the achievement, since it is trivial in amount.

On the face of it, this evidence of a risk taking preference in gambling linked to achievement motivation is something of a paradox. Persons who are highly motivated to achieve should *not* experience any sense of accomplishment, nor anticipate it, when as a matter of fact the outcome is a chance event, one for which they cannot feel personally responsible. Nor should failure to win be reacted to as a personal threat when the motive to avoid failure is strong. The fact that the high n Achievement group *did not* show a statistically significant preference for intermediate vs. extreme risk bets when compared with the chance expectation is consistent with this supposition that games of skill and games of chance should be reacted to differently. The further washing out of the slightly greater

preference for intermediate risk by the high n Achievement group when the amounts of money are increased suggests that the slight preference for intermediate risk with small amounts of money can be understood as a simple generalization of the tendency developed in competitive achievement situations requiring skill that is checked when the chance nature of the activity becomes more salient. Yet the low n Achievement group clearly avoided intermediate risk even when the amount of money was increased. This finding may mean that persons who have a long history of anxiety in competitive achievement activities fail to discriminate between games of skill and games of chance because the outcome more often seems beyond their control even in tasks requiring skill.

The present results regarding betting preferences must be considered little more than exploratory. We have already begun to follow them up. The small expected value (30¢) betting questionnaire described in this paper has been used in two small follow-up studies with male *S*s. When the results for all three studies are combined, it appears that among *S*s high in n Achievement ($N = 44$): 16% preferred 5/6, 14% preferred 4/6, 25% preferred 3/6, 7% preferred 2/6, and 11% preferred 1/6, while 27% showed no clearcut preference. Among *S*s low in n Achievement ($N = 36$): 33% preferred 5/6, 14% preferred 4/6, 11% preferred 3/6, 0% preferred 2/6, and 28% preferred 1/6, while 14% showed no clearcut preference. The chi square for intermediate (4/6, 3/6, 2/6) vs. extremes (5/6, 1/6) is 7.10 ($df = 1$, $p < .005$ in the predicted direction). In addition, "goodness of fit" tests show that the high n Achievement group's preference for intermediate vs. extreme risk is not different from what is expected by chance ($\chi^2 = .08$, $df = 1$, $p > .75$) while the low n Achievement group very strongly prefers the extreme risk bets ($\chi^2 = 12.39$, $df = 1$, $p < .001$). Preliminary data has also been collected on zero-expected value bets and on small negative-expected value bets (lose 30¢). In neither case, however, was any relationship found between n Achievement and a tendency to take intermediate vs. extreme risks.

As tenuous as these results presently appear, they are of considerable importance in light of Edwards' (1953) earlier report of probability preferences in gambling and the tendency of persons engaged in research related to probability-utility theory of decision scarcely to pay even lip service to the notion that anything but the value of money qua money affects the utility of the outcome in simple bets.

SUMMARY

This study deals with the effects of individual differences in strength of achievement motive on (*a*) goal setting inferred from choices individuals make among tasks varying in difficulty and (*b*) preferences for imaginary bets equated for expected monetary value but differing in probability of winning. The stimulus for this research is a risk taking model which predicts that, when the motive to approach success is stronger than the motive to avoid failure, tasks with intermediate probabilities of success will be preferred; but, when the motive to avoid failure is stronger, intermediate risk will be avoided and tasks with very high or very low probabilities of success will be preferred. Measures of n Achievement were obtained by content analysis of responses to the French Test of Insight. Persons with high n Achievement scores were assumed to have relatively stronger motives to approach success and those with low n Achievement scores relatively stronger motives to avoid failure.

Sixty-six college men played a shuffleboard game in which they were allowed to shoot from any of 15 lines varying in distance from the target. *S*s high in n Achievement initially shot from intermediate distances, 6 to 10 ft., significantly more often than *S*s low in n Achievement. However, when it became apparent after practice that the probability of success was .50 much closer to the target, *S*s high in n Achievement then took more shots than *S*s low in n Achievement from up close.

The high n Achievement group also preferred intermediate risk bets (4/6, 3/6, 2/6) over extreme risk bets (5/6, or 1/6) significantly more often than the low n Achievement group when the expected monetary value of the bet was small (30¢).

The results are consistent with the theory concerning motivational determinants of risk taking behavior and constitute evidence of the

possible effects of individual differences in strength of achievement motive on risk taking in games of chance as well as in tasks requiring skill.

REFERENCES

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372. (Also reprinted in Atkinson, 1958b)

ATKINSON, J. W. Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. (a)

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958. (b)

ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety as motives to approach success and to avoid failure. *J. abnorm. soc. Psychol.*, 1959, **58**, 52–63.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366. (Also reprinted in Atkinson, 1958b)

BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 467–476.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

EDWARDS, W. Probability preferences in gambling. *Amer. J. Psychol.*, 1953, **66**, 349–364.

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

EYSENCK, H. J., & HIMMELWEIT, H. T. An experimental study of the reactions of neurotics to experiences of success and failure. *J. gen. Psychol.*, 1946, **35**, 59–75.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In Atkinson, J. W. (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

LITWIN, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, Univer. of Michigan, 1958.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In Hunt, J. McV. (Ed.), *Personality and behavior disorders*, Vol. 1, Chap. 10. New York: Ronald, 1944.

MCCLELLAND, D. C. Some social consequences of achievement motivation. In Jones, M. R. (Ed.), *Nebraska symposium on motivation, 1955*. Lincoln: Univer. Nebraska Press, 1955.

MCCLELLAND, D. C. Risk-taking in children with high and low need for achievement. In Atkinson, J. W. (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **16**, 115–118.

MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–387.

MOULTON, R. W., RAPHELSON, A. C., KRISTOFFERSON, A. B., & ATKINSON, J. W. The achievement motive and perceptual sensitivity under two conditions of motive arousal. In Atkinson, J. W. (Ed.) *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

RAPHELSON, A. C. The relationship between imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18. (Also reprinted in Atkinson, 1958b)

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315–326.

VITZ, P. The relation of aspiration to need achievement, fear of failure, incentives, and expectancies. Unpublished honors thesis, Univer. of Michigan, 1957.

(Received October 8, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE INTEREST VALUE AND RELEVANCE OF FEAR AROUSING COMMUNICATIONS

LEONARD BERKOWITZ AND DONALD R. COTTINGHAM
University of Wisconsin

SEVERAL years ago Janis and Feshbach (1953) tested the "common sense" belief that attitudes are readily changed by fear arousing communications. As they point out, this notion, widely shared by health and safety educators as well as many propagandists, assumes that the frightened listener becomes highly motivated to accept the reassuring recommendations advocated by the communicator. However, contrary to this popular view, Janis and Feshbach showed that opinions presented through a strong fear arousing communication were less readily adopted by high school students than opinions presented in a talk making minimal use of fear.

The Yale investigators caution against overly hasty generalizations from their findings. Conceivably, they suggest, fear appeals may prove to be highly successful under certain conditions. The present study is an attempt to specify some of these conditions.

One possibility has to do with the interest value of the communication. A communicator may not succeed in changing his audience's opinions if his message lacks sufficient interest, even though he possesses other prerequisites for successful influence, such as high credibility (Hovland, Janis, & Kelley, 1953). People required to listen to or read an uninteresting message undoubtedly suffer some frustration, particularly if the message is long. As in the case of other frustration reactions, the bored audience might respond by "withdrawing from the field," i.e., not attending to the communication, or by directing hostility toward the communicator. Thus, it is likely that relatively high interest value is necessary but not sufficient for creating opinion changes through mass communications, either because low interest produces inattentiveness or because the frustrated audience, venting its aggression, disparagingly dismisses the entire communication.

This hypothesis can easily be reconciled with the results of the earlier study. There, even though the strong appeal was of somewhat greater interest to the *Ss* than the moderate and minimal fear appeals (though the former condition was least effective in changing attitudes), the majority of *Ss* in all conditions thought that "all or most" of the communication they received was interesting. The message, then, probably was above threshold in interest for most *Ss* in all conditions, permitting the operation of other attitude changing variables. However, other communications in other settings may be low in intrinsic interest for the audience. The dramatic nature of anxiety provoking material may heighten the interest value of a communication above the required threshold so that under certain conditions the strong fear appeal may successfully change opinions. In contrast, a minimal fear appeal, because less interesting, may produce little opinion change.

A second proposition concerning the success of a fear appeal assumes individual differences in the strength of emotional reactions to the communication. The importance of this assumption can be seen in the theoretical analyses made by Janis and Feshbach. To account for their results they raise the possibility of "defensive avoidance". "When fear is strongly aroused but is not fully relieved by the reassurances contained in a mass communication, the audience will become motivated to ignore or to minimize the importance of the threat" (1953, p. 90). In other words, the audience defensively avoids accepting the beliefs advocated by the communicator. However, in a later re-examination of their data (1954), the Yale psychologists also point out that some *Ss* will become more fearful than others as a result of the threatening appeal. Contending that people with a chronically high anxiety level are relatively vulnerable to fear arousal, they show that "persons who manifest symptoms of acute anxiety in their everyday lives" were consistently less influenced by the strong appeal than low anxiety *Ss*. The former *Ss*, presumably because of

their stronger fear, exhibited a stronger defensive-avoidance reaction.

In this case, then, the degree of fear aroused by the threatening communication seems to be a direct function of *S*s' characteristic level of anxiety. Extrapolating from Miller's (1948) conflict model relating the strength of the individual's avoidance tendencies to his psychological distance from a desired yet feared goal, we can also propose that the magnitude of the avoidance-fear reaction increases with the relevance of the message to the needs of the audience members. Those *S*s for whom the message is highly relevant obviously are closer to making the appropriate goal responses called for by the communication, i.e., they are closer to using the advocated beliefs than the *S*s for whom the material is less relevant. Therefore, since *S*'s fear and the resulting tendency-to-avoid response increases with his approach to the feared situation in which the response would be elicited, a stronger defensive avoidance tendency should be produced by a relevant than by an irrelevant communication.

Combining the two lines of reasoning presented above, we predict: (*a*) To the extent that the minimal fear appeal is relatively uninteresting to the audience, there should be little, if any, opinion change in the *S*s receiving this communication. (*b*) Assuming that the strong appeal is of sufficient interest, the less relevant the material to the *S*s in this condition the greater the opinion change in the direction advocated.

METHOD

Procedure

Two separate experiments were conducted employing different *E*s and different "communicators" but using the same communications in the same procedure.[1] In each case, the *S*s, all volunteers from introductory psychology courses, were recruited ostensibly as part of a testing program designed to evaluate the effectiveness of various lecture techniques. When they came to the assigned rooms, in groups of at least 18 members, the *S*s in the experimental conditions listened to a tape recorded lecture on automobile safety belts, and then answered a brief questionnaire indicating their opinions of the lecture, the speaker, and the desirability of having safety belts. A control group responded only to the safety belt items without hearing the tape recording.

Two lecture conditions were employed in each study: minimal and strong fear arousal. In the former condition, after a 370-word introduction which was common to both lectures, the speaker went on to give a 350-word emotionally neutral argument elaborating the advantages of automobile safety belts and relying mainly upon statistical data together with one cartoon-like slide. This special material, read by the same speaker, not only was longer in the strong fear condition (510 words) but also was more personal and dramatic, and made use of 10 slides, most of which depicted gruesome car accidents. Essentially the same arguments were used in both lectures. The lectures remained unchanged from one study to the next, although a different male speaker was employed in each experiment.

*S*s' attitudes towards safety belts, as assessed by an 18-item Likert scale included in a questionnaire administered immediately after the experimental lecture (or the day following the experimental sessions but without a lecture in the case of the control *S*s) were then compared with their attitude scores on the same items approximately one month earlier. For this pretest the *S*s, who had volunteered to serve in a study of "the psychology of normal people", had been brought together in relatively large groups and given a 123-item questionnaire dealing with such topics as national tariffs, movies, Jews, and the punishment of criminals, as well as automobile safety belts. The *S*s were told that *E* was interested in determining how college students felt about these issues. Several weeks later a different *E* contacted each of the *S*s, informed him (or her) of the lecture effectiveness study, and solicited his participation in the experiment. No mention was made of the earlier attitude testing.

Questionnaire Measures

Two items in the postexperiment questionnaire were designed to test the effect of the fear arousal manipulations. The first of these items reads, "I would classify the presentation as . . ." with the responses ranging in five steps from "very pleasant" through "pleasant," "nonstimulative," and "unpleasant" to "very unpleasant." The second item was worded, "The presentation of the material was . . .," with alternatives "*a*) boring, *b*) of indifferent quality, *c*) thought provoking and *d*) disturbing."

The responses to the items were combined in Experiment I by weighting each alternative in terms of the amount of discomfort indicated and adding the two scores. The higher the score, then, the higher the admitted tension evoked by the communication. However, in Experiment II the items were kept separate for two reasons. First, answers to the items exhibited a somewhat different pattern in the second study. Fewer people in the latter investigation described themselves as very upset by the lecture, particularly in response to the second question, while a greater proportion of the *S*s indicated the material was uninteresting to them. Second, the greater fre-

[1] Experiment I was carried out as part of a course in experimental social psychology conducted by the senior author. Each member of the class participated in the research in some way. Experiment II was carried out by D. R. Cottingham under the supervision of L. Berkowitz.

quency of responses indicative of boredom in Experiment II made it possible to test the effects of communication interest-value upon opinion change.

The relevance of the lectures to the *Ss* also was assessed somewhat differently in the two investigations. In the first experiment only two levels of relevance were established, defined solely in terms of frequency of using a car: once a week or less vs. at least several times a week. Car ownership as well as frequency of car usage was taken into account in determining three levels of relevance in Experiment II. Those who indicated that they both owned an automobile and drove it at least several times a week were placed in the High Relevance group. The Medium Relevance group either owned a car and drove it less frequently, or did not own an automobile but rode in one at least several times a week. The Low Relevance *Ss* neither owned a car nor rode in one more often than on weekends or once a week.

The scale to measure attitude towards safety belts included statements ranging from "automobile safety belts should be required by law" to items maintaining that their value had not yet been proven. In both experiments the *Ss* indicated whether or not they agreed with the item. For convenience, we will refer to a change from the initial attitude score to the post-experiment score as a change in the number of items accepted, although change in the number of negative items rejected is also involved. The mean change in each condition was toward the advocated beliefs.

To summarize, the experimental procedure was the same in Experiments I and II, although a different speaker was used in each. The major difference between the two studies lies in the classification of relevance levels within the strong, minimal, and control conditions. The *Ss* in Experiment I were divided into two relevance levels in terms of their frequency of car usage, while the *Ss* in the second study were divided into three such levels on the basis of both car ownership and usage. The main dependent variable, a Likert scale of attitudes towards automobile safety belts, was the same in both investigations, but there are some differences in the amount of supplementary information available. Where Experiment I affords only a single composite index of discomfort induced by the communication, Experiment II provides information concerning both the tension felt by the *Ss* and their degree of interest in the communication.

RESULTS

Experiment I

Six groups of 13 *Ss* each were formed by dividing each of the experimental conditions (the strong and minimal fear appeals conditions) and the control condition into the two relevance levels. Table 1 shows the mean discomfort score induced by the communications in each of the resulting four experimental groups. It is clear that the manipulation of fear arousal was successful. The *Ss* in the strong fear arousal condition admitted to significantly greater feelings of tension than the *Ss* receiving the minimal fear appeal ($t = 2.51$; $p = .02$).[2]

TABLE 1
MEAN DISCOMFORT SCORE IN EACH OF THE EXPERIMENTAL GROUPS
(Experiment I)

Condition	Mean Discomfort Score	Significance of Differences
Strong Fear Arousal		
High Relevance	6.54	H.R. vs. L.R.:
Low Relevance	5.62	$p = .05$
Minimal Fear Arousal		
High Relevance	4.92	H.R. vs. L.R.:
Low Relevance	5.08	NS

Note.—Strong vs. Minimal: $p = .02$.

The next table presents the mean opinion change in the direction advocated by the communication within each of the six conditions. As is readily apparent, there are no reliable differences among the experimental and control conditions when all the *Ss* are combined across relevance levels. However, when we consider the relevance of the message, our theoretical expectations are confirmed. The Low Relevance *Ss* in the strong fear condition exhibited a significantly greater change in the direction advocated than either the High Relevance *Ss* in this condition or the similar Low Relevance *Ss* in the control condition. (These differences, the only ones reliable, are significant at $p < .05$.)

The table also suggests that the strong fear appeal was somewhat (but not significantly) more effective than the minimal fear appeal for the Low Relevance *Ss*. It may be that the former communication possessed greater interest value for the audience, permitting the members to be convinced by the logic of the speaker's message, while the same arguments could not be effective when they were embedded in an uninteresting communication.[3]

[2] The error terms in the two-tailed t tests used for the condition comparisons in this paper are based upon the residual mean squares obtained from preliminary analyses of variance.

[3] The relatively great increase in favorableness of attitudes towards safety belts in the control conditions suggests that many *Ss* came to believe automobile safety belts were desirable after being stimulated to think about this issue.

TABLE 2
MEAN CHANGE IN ATTITUDE TOWARD SAFETY BELTS: EXPERIMENT I
(N = 13 in each condition)

	Control	Minimal Fear	Strong Fear
High Relevance			
Initial attitude	4.19	3.81	4.04
Mean change	3.29	3.48	3.14
Low Relevance			
Initial attitude	4.14	4.34	4.40
Mean change	2.90	3.83	4.25

TABLE 3
NUMBER OF *Ss* GIVING THE FOLLOWING REACTIONS TO THE ITEM, "I WOULD CLASSIFY THE PRESENTATION AS . . .", IN EXPERIMENT II

Condition	Pleasant	Non-stimulative	Unpleasant
Strong fear appeal	16	2	8
Minimal fear appeal	7	12	1

Note.—Chi square (employing Yates' correction) = 14.26; $p < .001$.

Experiment II

The second experiment, using 64 *Ss*, was designed to test the foregoing interpretations. One item on the postquestionnaire, identical to that employed in the earlier study, indicates that the strong fear appeal again created a somewhat stronger feeling of discomfort in the audience than the minimal fear appeal. The data are presented in Table 3. The speaker employed in this study apparently was less effective than the speaker used in the earlier investigation; although the script was identical, only about 31% of the *Ss* in the strong fear condition (as compared with only one *S*—5%—in the minimal fear condition) now admitted to some feelings of unease. The chi square for this difference, based upon only those *Ss* classifying the presentation as "unpleasant" and employing Yates' correction, is 4.00 (p = .05).

The table also indicates that the people in the strong fear condition were more likely to call the presentation "pleasant," and were less likely to be "indifferent" to it, than the *Ss* in the minimal fear condition. The chi square for the former reaction is 3.52 ($p < .10$), and for the latter is 7.14 ($p < .01$). Taken together, these differences suggest that the strong appeal was more interesting to the audience than the minimal fear arousing lecture. Further evidence is provided in the analysis of the second questionnaire item summarized in Table 4. Forty per cent of the *Ss* in the strong fear condition indicated on this item that the lecture had low interest value, while this figure rose to 75% in the minimal fear condition.

According to the present hypotheses, then, opinion changes are more likely to occur through the relatively interesting strong fear arousing lecture than in the more boring minimal fear arousing condition. This expectation is confirmed by the data presented in Table 5.

As can be seen in Table 5, the over all pattern of results is similar to the findings obtained in Experiment I. First, the *Ss* receiving the fairly uninteresting minimal fear appeal were not more likely than the control *Ss*, to change in the direction advocated, and there was no significant relationship between the relevance of the communication and the amount of opinion change for the *Ss* in these conditions. Second, consistent with Prediction *b*, in the case of those *Ss* receiving the more interesting strong fear-arousing communication, the less relevant the material the greater the opinion change in the direction advocated. In the strong fear condition the Low Relevance *Ss* have a greater mean change score (in terms of number of items accepted) than either the High or Medium Relevance *Ss* (though the latter t does not attain significance; t with the High Relevance *Ss* is 3.23; p = .01), while the Medium Relevance *Ss* exhibit a reliably greater change than the High Relevance *Ss* (t = 2.28; p = 0.05). Moreover, the Low Relevance *Ss* receiving the strong fear appeal have a significantly greater mean change score than any of the

TABLE 4
NUMBER OF *Ss* CLASSIFYING THE EXPERIMENT II PRESENTATION AS LOW IN INTEREST-VALUE ("BORING" OR "OF INDIFFERENT QUALITY") OR NOT LOW IN INTEREST-VALUE ("THOUGHT PROVOKING")

Condition	Low Interest	Not Low Interest
Strong fear appeal	10	15
Minimal fear appeal	15	5

Note.—Chi square = 5.99; p = .05.

TABLE 5
MEAN CHANGE IN ATTITUDE TOWARD SAFETY BELTS: EXPERIMENT II

	Control	Minimal Fear	Strong Fear
High Relevance			
N	6	7	11
Initial attitude	6.0	6.4	6.3
Change score	2.2	1.7	2.7
Medium Relevance			
N	6	6	9
Initial attitude	5.7	5.5	7.2
Change score	1.2	3.3	6.2
Low Relevance			
N	6	7	6
Initial attitude	7.3	6.7	6.8
Change score	2.2	2.7	8.3

groups in the minimal fear and control conditions ($p < .01$ in every case with the exception of the .05 level difference with the Medium Relevance *Ss* in the minimal fear condition). None of the other differences is significant.

Other Effects of Fear Arousal

In keeping with the Yale approach to studies of opinion change (Hovland, Janis & Kelley, 1953), Janis and Feshbach asked whether their condition differences could be attributed to differences in the learning of the material presented and/or differences in *Ss'* attitudes toward the communicator. According to the first of these possibilities, the defensive avoidance tendency provoked by the strong fear arousal could have produced inattentiveness to the communication. However, they obtained no evidence for such a process; *Ss* in the strong fear condition learned as much of the communicated information as those in the less fearful conditions. The present theoretical formulation also necessitates obtaining indices of the degree of attention in the various communication conditions. We have proposed that the frustrating experience of listening to a boring speech may result in greater inattentiveness to the lecture. Again, however, there is no evidence of differences in the learning of the lecture material. On an information test included in the postexperiment questionnaire (in Experiment II), *Ss* in both experimental conditions at each relevance level were correct, on the average, on approximately 7 of the 11 items.

The second possibility mentioned and then discarded by Janis and Feshbach was that the fear-arousing communication provoked hostility toward the communicator which in turn interfered with the acceptance of his argument. We also have suggested that aggression towards the communicator may be a factor in producing little opinion change, but that frustration may develop during the relatively uninteresting minimal fear arousing lecture. To test this, an item in the post-experiment questionnaire used in the second study asked *S* to indicate whether the lecturer sounded very pleasant, pleasant, neutral, unpleasant, or very unpleasant. We assume that *Ss* who feel angry towards the speaker would be more likely to judge him as unpleasant. Contrary to our expectations, 31% of the 26 *Ss* in the strong fear condition rated the speaker as unpleasant, while only 10% of the 20 *Ss* in the minimal fear condition made this judgment. However, the chi square value for this difference is only 1.69, employing Yates' correction, which is nonsignificant.

It would appear, then, that the relative ineffectiveness of the minimal fear arousing lecture cannot be traced either to failure to attend to the communication or aggression toward the communicator. Any hostility instigated by the presumably frustrating experience of listening to the uninteresting lecture probably was directed at the communication (as indicated by the disparaging evaluations of the presentation in the minimal fear condition) and/or to the experimenters.

DISCUSSION

Little direct evidence has been presented so far for the process hypothesized to account for the relatively great opinion change in the Low Relevance subgroup. Contrary to the present formulation, these *Ss* within the strong appeal condition may have accepted the communication, not because they had a low avoidance tendency, but because, lacking driving experience, they were more susceptible to authoritative persuasion in this area. Clearly, they were not generally more susceptible to arguments favoring safety belts; the Low Relevance *Ss* in the minimal appeal condition did not tend to adopt the advocated beliefs. However, as we already have suggested, the

low level of opinion change under minimal fear appeal may have been the result of an uninteresting communication, while the strong appeal may have been more convincing. Given this necessary interest value, the inexperienced Low Relevance *S*s could have been more susceptible to an expert's persuasion.

Nevertheless, there is reason to believe that the different relevance groups were quite similar in their driving experience. With the exception of two people at each relevance level, all *S*s in the second experiment indicated, in response to a questionnaire item, that they had been driving more than three years. The *S*s, it may be noted, were on the average between 19 and 20 years old.

But what is really needed to buttress the hypothesis that relevance increases the defensive-avoidance reaction to fear-arousing communications is evidence that anxiety (or discomfort) varies directly with relevance. Some of the data suggest that this is indeed the case. The results for Experiment I, presented in Table 1, indicate that the High Relevance *S*s receiving the strong fear appeal were significantly more upset by the communication than the *S*s for whom the message was less pertinent, while there was no difference in admitted tension between the two relevance groups in the minimal fear condition. This significant difference in the strong fear condition was not obtained in Experiment II because of the small numbers of cases, but here too there are indications of greater anxiety in the more relevant groups. Findings reported earlier suggest that evaluations of the lecturer in this study tend to bear a direct relationship to the discomfort created by his communication; he was more likely to be judged as unpleasant in the strong appeal than in the minimal appeal conditions. Closer inspection of the data reveals that the lowest ratings of the speaker were given by the High and Medium Relevance *S*s within the strong appeal treatment. Three of the former called him unpleasant, five in the latter subgroup did so, as compared with none of the Low Relevance *S*s. Employing Fisher's exact probability test, the difference between the High and Low Relevance *S*s is nonsignificant ($p = .24$), but there is a significant difference ($p = .04$) between the Medium and Low Relevance *S*s. Apparently, the fear-arousing speaker was somewhat more unpleasant to the High and Medium Relevance *S*s than to the Low Relevance *S*s.

In both experiments, then, there are indications that relevant fear-arousing communications produce greater tension in the audience than less relevant messages. If this should be the case, the present results are consistent with an hypothesis advanced by Janis and Feshbach (1954), who propose that a fear appeal is especially likely to be accepted when a relatively weak anxiety drive is aroused. Further research is necessary, however, before we can conclude with certainty that the High Relevance *S*s are less likely to accept the fear arousing appeal because of a tendency toward defensive avoidance.

SUMMARY

An experiment by Janis and Feshbach (1953) indicated that opinion change in the direction advocated by a communication is less likely to occur when the communication arouses strong fear than when it arouses less fear. In explanation, they suggested that fear arousal caused the *S*s defensively to avoid accepting the communicator's conclusions. The two experiments reported here were designed to show that a strong fear appeal could be more convincing than a weak one when (*a*) the communication is low in interest value and the dramatic nature of the "strong" communication makes it considerably more interesting than the "weak" communication, and (*b*) the communication is of low relevance to the actions of the audience.

The two experiments, employing college students who listened either under strong fear or weak fear conditions, to a tape recorded lecture advocating the use of automobile safety belts, agreed in confirming these expectations. There was little opinion change with the relatively uninteresting minimal fear lecture, while the degree of opinion change produced by the more interesting strong fear lecture was inversely related to the relevance of the material to the *S*s. Miller's hypothesis concerning the slope of the avoidance gradient was cited in accounting for the findings with the relevance dimension.

REFERENCES

HOVLAND, C., JANIS, I., & KELLEY, H., *Communication and persuasion*, New Haven: Yale Univer. Press, 1953.

JANIS, I., & FESHBACH, S. Effects of fear-arousing communications. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.

JANIS, I., & FESHBACH, S. Personality differences associated with responsiveness to fear-arousing communications. *J. Pers.*, 1954, **23**, 154–166.

MILLER, N. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

(Received October 15, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

EVALUATIONAL REACTIONS TO SPOKEN LANGUAGES[1]

W. E. LAMBERT, R. C. HODGSON,[2] R. C. GARDNER, AND S. FILLENBAUM[3]
McGill University

SPOKEN language is an identifying feature of members of a[4] national or cultural group and any listener's attitude toward members of a particular group should generalize to the language they use. From this viewpoint, evaluational reactions to a spoken language should be similar to those prompted by interaction with individuals who are perceived as members of the group that uses it, but because the use of the language is one aspect of behavior common to a variety of individuals, hearing the language is likely to arouse mainly generalized or stereotyped characteristics of the group. Thus, when one hears a radio broadcast of an international meeting and encounters passages of a foreign language, one's evaluational reactions to the communication are attributable, in part, to the language used and likely reflect generalized attitudinal reactions to the group that uses it.

The purpose of the present investigation was to determine the significance spoken language has for listeners by analyzing their evaluational reactions to English and French. Since we were interested in reactions that are attributable primarily to the language itself, we attempted to minimize the effects of both the voice of the speaker and his message by employing bilingual speakers reading the same message in two languages. In view of previous studies (see Licklider & Miller, 1951, p. 1070f) which have shown that evaluations of personality based solely upon voice have little or no reliability, it was predicted that the differences in the favorableness of any *S*'s evaluations of the French and English guises of speakers would reflect his attitude toward members of his own and members of the other language group. The study was carried out with *S*s living in Montreal, a community whose history centers largely in a French-English schism which is perhaps as socially significant for residents of the Province of Quebec as that between the North and the South is for Southerners in the United States.

[1] This research was supported by the Canadian Defence Research Board, Grant Number D77-94-01-10, and by a Ford Foundation grant to D. O. Hebb. We are grateful to Henri Barik for his computational assistance.

[2] Now with the Aluminum Company of Canada.

[3] Now at the University of North Carolina.

METHOD

Procedure

A 2½ min. passage of French prose of a philosophical nature was translated into fluent English and tape recordings were made of the voices of four male bilinguals each of whom read both French and English versions of the passage. Recordings were also made of the voices of two other men, one reading the passage in English, the other in French. There were, then, 10 taped voices, four of which were "matched," each speaker using both languages, and two used as "filler" voices and for practice. The 10 voices were presented to *S*s in alternating French-English order starting with the two filler voices and allowing the maximum possible interval between successive presentations of the English and French guises of any speaker. Evaluational reactions to the matched voices only were examined.

The study was introduced as an experimental investigation of the extent to which peoples' judgments about a speaker are determined by his voice and *S*s were reminded of the common phenomenon of having a voice on the radio or telephone summon up a picture in the listener's mind of the person speaking. *S*s were *not* told that they were going to hear some of the voices twice, but rather that they would hear 10 recorded male voices, all reading the same passage, five in French and five in English. The two languages, it was said, were being used to give greater scope to the experiment. In this regard, *S*s were given copies of the French and English versions of the text they were to hear so that they could be acquainted with the message and the languages and be better prepared to pay attention only to the voices of the speakers. There was no indication that any *S* became aware of the fact that bilingual speakers were used.

The *S*s were also given a response sheet for each voice which directed them to rate each of 14 traits on 6 point scales ranging from "very little" (*fort peu*, for French *S*s) to "very much" (*beaucoup*). The traits used were: height (*taille*), good looks (*attrait physique*), leadership (*apte à diriger*), sense of humour (*sens de l'humour*), intelligence (*intelligence*), religiousness (*pieux*), self-confidence (*confiance en soi*), dependability (*digne de confiance*)[4], entertainingness (*jovialité*), kindness (*bonté*), ambition (*ambition*), sociability (*sociabilité*), character (*caractère*) and general likeability (*est-ce qu'il est sympathique*?). Each voice was played once and *S*s rated that voice on the scales both

[4] *Digne de confiance* translates to both "dependable" and "trustworthy."

while it was playing and during the 90 sec. interval between voices. At the bottom of each response sheet *Ss* were asked to indicate in what occupation this man would likely be found.

Attached to the back of the 10 response sheets, facing down and covered by a blank sheet, were several questionnaires to be completed. This packaging of questionnaires and response sheets discouraged *Ss* from looking ahead in the booklet and also permitted them to answer all items anonymously. *Ss* were asked only their age, place of birth and religious affiliation. After the ratings were obtained on all voices, *Ss* were directed to turn to the first questionnaire.

Importance ranks of traits. The list of traits used to describe each voice was presented again and *Ss* were asked to rank the traits in terms of their desirableness in friends. It was felt that this information would offer a more personally relevant set of traits for analysis.

Prejudice scale. In view of its wide use as an index of generalized prejudice and ethnocentrism, we incorporated Forms 40 and 45 of the California F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) as one of the independent measures of attitude toward outgroups.

Attitude scales. *Ss* were asked to complete 14 incomplete sentences designed to elicit attitudes toward both their own and the other language group. Examples are: English Canadians think . . .; Children of French Canadian parents . . . ; The more I get to know French Canadians . . . ; French Canadians . . . ; English Canadians . . . ; and The more I get to know English Canadians. . . . The items were translated into French for the French speaking *Ss*. Two separate scores were assigned each *S* on the basis of his completions: an "attitude toward English group" score and an "attitude toward French group" score each comprising the number of favorable statements made minus the number of unfavorable statements made about each group. Two judges independently scored the items for each sample and the interjudge reliabilities are, for the English sample, $r = .69$ and $r = .86$ for attitudes toward English and French respectively and for the French sample $r = .75$, and $r = .80$.

Preference scale. *Ss* indicated their preference for English or French Canadians as: marital partners, friends, neighbours, tenants, colleagues, and political candidates, by ranking which they would rather have for each case. Preference scores were the sum of differences in ranks given to French and English Canadians for all items.

Degree of bilingualism. We reasoned that differences in evaluational reactions to the English and French guises would be less pronounced for those *Ss* who had experience with both languages and presumably both linguistic communities as well. *Ss* checked their degree of facility in the use of the other group's language by indicating how well they could speak, read, and write, and whether they could do so: not at all, a little, fairly well, or fluently. An index of bilingualism was computed for each *S* by assigning zero to "not at all" entries, and 3 to "fluently." We doubled the score for ability to speak the language on the assumption that speaking experience was most important for the purpose of the study. A maximum score would be 12. A bimodal distribution of scores was obtained from both samples and both were split at the same point so that *Ss* with scores 0 to 7 are categorized as being less bilingual and those with scores 8 to 12 as being more bilingual.

Subjects

The English speaking sample comprised 64 students, taking the first course in Psychology at McGill University, who volunteered to participate. The average age for the group was 18.8 years and both sexes were approximately equally represented.[5] The introduction and procedures were presented to the group by the first two authors in English. All *Ss* used English as their school and primary social language, and all but nine who were born in Europe learned English as their first language.

The French speaking sample was made up of 66 male students in their final year at a classical French *collège* in Montreal who were at approximately the same educational level as the English sample, with an average age for the group of 18.2 years. The study was presented to the students by a professor of the *collège* and the first author, in French. The experiment took the place of a regular lecture. All but one of these *Ss* were born in Canada and all but four in the province of Quebec. All used French as their home, school, and primary social language. The faculty and students speak grammatically correct French, of course, but in general they have a distinctive French Canadian accent.

Speakers

Much care was taken to select appropriate speakers who could take on both French and English guises (in the sense of habitual manners of speaking). We chose from acquaintances in the English speaking community three bilinguals (Bla, Cou, and Leo) who spoke faultless English and yet were trained in French schools in Canada through the graduate level and used French in their homes, or, in the case of Leo, in social or work situations. Several recordings had to be made of each speaker to have errorless readings and in the case of Bla to perfect the pronunciation of a few English words. The final recordings of all four speakers were judged to be spoken in perfect French and English by three bilingual judges. The judges agreed that speakers Bla and Cou spoke with French Canadian accents and that Leo spoke with a marked French Canadian accent characteristic of those who work "in the bush" where, in fact, his style of pronunciation was acquired. The

[5] Thirty-two *Ss* indicated that their religion was Jewish, and, to determine their comparability with the rest of the sample, a separate analysis was carried out for Jewish and non-Jewish *Ss*. For both subgroups, correlations were calculated between each of the four attitude measures and the comparative favorableness of evaluations of the English and French guises considering all traits as well as those most desirable for friends. The comparison of correlations indicated that there were no statistically significant differences between the two subgroups at any point and therefore the Jewish and non-Jewish *Ss* were combined.

TABLE 1

t VALUES FOR SIGNIFICANCE OF DIFFERENCES IN EVALUATIONS OF ENGLISH AND FRENCH GUISES OF SPEAKERS

Trait	English *Ss*' Judgment of Speakers					French *Ss*' Judgments of Speakers				
	Cou	Bla	Leo	Tri	Over-all	Cou	Bla	Leo	Tri	Over-all
Height	8.63**	4.67**	4.05**	2.59*	8.83**	5.50**	2.32*	2.41*	1.04	4.58**
Good Looks	7.16**	4.10**	2.66**	−.90	5.78**	7.50**	6.00**	7.26**	1.73	9.77**
Leadership	1.72	.69	1.41	−1.82	1.20	4.05**	9.32**	12.10**	.82	11.06**
Sense of Humor	−.58	−.14	2.25*	−3.48**	−2.16*	2.67**	.58	.64	−.30	1.40
Intelligence	1.44	.11	2.25*	1.98*	2.40*	4.54**	9.17**	11.73**	1.00	10.22**
Religiousness	−.67	−.33	−.26	1.81	.27	−1.79	−3.11**	−1.60	−1.14	−2.94**
Self-confidence	−.50	−.47	1.13	−1.33	−.69	5.37**	7.56**	10.00**	−2.00*	8.28**
Dependability	1.61	−.15	3.51**	2.74**	3.11**	2.89**	4.36**	7.50**	1.89	6.44**
Entertainingness	1.77	1.67	.00	−4.44**	−.84	.44	−.27	−.81	.22	−.21
Kindness	2.40*	2.47*	1.67	.74	3.37**	.10	−4.12**	−2.65*	−.74	−2.96**
Ambition	.45	2.43*	2.33*	.84	2.83**	3.06**	4.64**	7.53**	−1.53	5.89**
Sociability	.24	−.16	−.17	−4.13**	−1.44	3.44**	2.10*	3.47**	.29	3.72**
Character	3.07**	1.00	3.17**	.12	3.08**	3.95**	6.81**	8.68**	.31	8.51**
Likability	2.38*	.26	1.38	−1.21	1.22	1.83	.80	2.05*	.79	2.00*

Note.—Positive entries indicate that English guises are evaluated more favorably than French, and minus entries indicate a more favorable evaluation for French guises. This is so for both French and English *Ss*' judgments.

* Significant at .05 level.

** Significant at .01 level, two-tailed tests.

fourth speaker, Tri, spoke French with an accent that was judged as indistinguishable from that used in France. It was felt that these variations in accent and style would not only make the task more interesting for *Ss* but would also lead to new aspects of the problem for further study, e.g., the reactions of both samples to Parisian French (Tri) and to caricatured French Canadian French (Leo). However, since the present study was not designed to deal comprehensively with variations in accent, attention will be mainly given to over-all comparisons of French and English languages until the final section where accent differences will be discussed.

RESULTS

Favorableness of reactions: English Ss' *judgements.* For each *S* on each of the 14 traits the difference (D) between the evaluations of the English and French guises of each speaker was noted, with account taken of the direction of the difference. These D values were then summed over the 4 speakers and over *Ss* and the departure of the D scores from zero was examined. As can be seen in the over-all column of Table 1, the English *Ss* evaluated the following seven traits significantly more favorably for the English than for the French guises: Height, Good looks, Intelligence, Dependability, Kindness, Ambition, and Character.[6]

[6] Although a rating toward the upper end of the scale does not necessarily indicate favorableness in the cases of "height" or "ambition," the term "favorableness" is a meaningful description for the majority of the judgments.

There were no significant differences in evaluations of Leadership, Religiousness, Self-confidence, Entertainingness, Sociability, and Likability. The French guises were evaluated significantly more favorably on Sense of Humor, but this finding is not clearcut since for one speaker the English guise was evaluated significantly more favorably on this trait. There was considerable consistency in the data for three of the four voices but the fourth (Tri) differed from the others on a number of traits.

Ss had ranked the 14 traits for their desirability in friends. For each *S* two ΣD scores were obtained, one for the three traits judged most desirable and one for the three traits judged least desirable, by noting the difference in evaluations of the English and French guises for each set of traits. English *Ss* evaluated the English guises significantly more favorably both for desirable traits ($t_{\text{desir}} = 2.13$) and the less desirable traits ($t_{\text{undes}} = 4.33$). There was no difference between the most and least desirable traits with regard to this preference ($t_{\text{diff}} = 1.72$).

Favorableness of reactions: French Ss' *judgments.* A similar analysis was carried out for the French *Ss*. They evaluated the English guises of the speakers significantly *more* favorably than the French guises for the following 10 traits: Height, Good looks, Leader-

TABLE 2

t VALUES FOR SIGNIFICANCE OF MEAN DIFFERENCES IN EVALUATIONS, ENGLISH VS. FRENCH JUDGES

Trait	English Guises	French Guises
Height	3.76**	0.30
Good Looks	1.78	4.59**
Leadership	−2.55*	6.19**
Sense of Humor	0.62	3.10**
Intelligence	−2.87**	2.82**
Religiousness	2.44*	−0.15
Self-confidence	−3.02**	4.85**
Dependability	−1.32	1.97*
Entertainingness	−0.52	−0.42
Kindness	1.82	−3.18**
Ambition	−0.57	2.28*
Sociability	−1.87	2.14*
Character	−1.22	3.60**
Likability	−1.34	−0.31

Note.—Positive entries indicate that English judges evaluated a particular trait more favorably on the average than did French judges; minus entries indicate more favorable evaluation for French than English judges.

* indicates a *t* is significant at the .05 level, on two-tailed tests.

** indicates a *t* is significant at the .01 level, on two-tailed tests.

ship, Intelligence, Self-confidence, Dependability, Ambition, Sociability, Character, and Likability (Table 1). There was no difference in evaluations with regard to Sense of Humor and Entertainingness, and the French guises were evaluated significantly more favorably on Religiousness and Kindness. The results for three of the four speakers were consistent and again speaker Tri was the exception. The English guises were evaluated significantly more favorably both for the three most desirable traits ($t_{87} = 7.17$) and for the three least desirable traits ($t = 7.58$) and, as in the case of the English *S*s, there was no difference between the most and least desirable traits ($t < 1$).

Favorableness of reactions: English Ss' vs. French Ss' judgments. The evaluations of the English guises given by the English judges were compared with the evaluations of the English guises given by the French judges, and a similar analysis was carried out for the evaluations of the French guises by the two groups of *S*s. Table 2 shows that for the English guises the English *S*s evaluated Height and Religiousness significantly more favorably than did the French *S*s. There were no differences between English and French *S*s with regard to: Good Looks, Sense of Humor, Dependability, Entertainingness, Kindness, Ambition, Sociability, Character, and Likability, and the French *S*s evaluated the English guises significantly more favorably than did the English S_o on Leadership, Intelligence, and Self-confidence. The comparisons with respect to the French guises are also given in Table 2. It can be seen that the English *S*s evaluated the French guises significantly *more* favorably than did the French *S*s on the following nine traits: Good looks, Leadership, Sense of Humor, Intelligence, Self-confidence, Dependability, Ambition, Sociability, and Character. There were no differences in evaluations for: Height, Religion, Entertainingness, and Likability, and the French *S*s evaluated the French guises significantly more favorably than did the English *S*s on Kindness.

Attitudinal measures. The correlations between the various measures of attitude and the comparative favorableness of evaluations of guises are given in Table 3 for the English *S*s. It is striking that none of the attitude measures correlate with the degree of favorableness of reactions to English and French guises, with the exception of attitudes to their own group as reflected in responses to the incomplete sentences, a measure that just meets significance requirements.

Although the pattern of correlations is also low for the French *S*s, the comparative favorableness of evaluations of guises for this sample is somewhat more clearly related to both attitudes toward and preference for own group. It should be noted that the French *S*s on the average clearly showed more favorableness to English than French guises, in fact only 14% of the *S*s showed the expected trend of more favorable reactions to French over English versions. This was not the case for either the preference or attitude measures where the majority of *S*s (85% and 62%, respectively) preferred and showed more favorable attitudes toward the French over the English group. The significant correlations in Table 4 therefore indicate that the more the French *S*s show preference for and favorable attitudes toward their own group the less they overrate English guises.

The relation of bilingualism to other variables. *S*s categorized as having a high degree of bilingualism were compared with those having little experience or skill in the other group's language on all attitudinal variables as well

TABLE 3

RELATION OF ATTITUDINAL MEASURES TO OTHER VARIABLES; ENGLISH SAMPLE

	1	2	3	4	5	6
Prejudice (F scale)	X	.29*	−.22	−.01	.20	.19
Preference (Bogardus)		X	−.40**	.17	.13	.13
Attitudes to French (incomplete sentences)			X	−.18	−.04	−.08
Attitudes to English (incomplete sentences)				X	.26*	.25
Favorableness of evaluations (English over French; all traits)					X	.77**
Favorableness of evaluations (English over French; desirable traits)						X

Note.—The *N*'s vary around 55 for all entries except Variable 6 where they drop to around 50.

* Significant at .05 level.

** Significant at .01 level.

TABLE 4

RELATION OF ATTITUDINAL MEASURES TO OTHER VARIABLES; FRENCH SAMPLE

Variables	1	2	3	4	5	6
Prejudice (F scale)	X	.17	.03	−.07	−.05	−.11
Preference (Bogardus)		X	.38**	−.45**	.26*	.27
Attitudes to French (incomplete sentences)			X	−.43**	.25*	.45**
Attitudes to English (incomplete sentences)				X	−.21	−.20
Favorableness of Evaluations (French over English; all traits)					X	.72**
Favorableness of Evaluations (French over English; desirable traits)						X

Note.—The *N*'s vary around 55 for all entries except Variable 6 where they drop to around 35.

Since the French *S*s in general showed more favorableness to English than French guises, Variables 5 and 6 mainly represent degrees of nonfavorableness to English guises.

* Significant at .05 level.

** Significant at .01 level.

as the comparative favorableness of evaluations to the spoken languages. For the English sample, the more bilingual subgroup had a reliably lower mean F score than did the less bilingual group, $t_{60} = 2.20$, a finding consistent with another study (Gardner & Lambert, 1958) on the relation of bilingualism to attitudes. The two bilingual groups did not differ reliably on any other variable, however. For the French sample, the only significant difference between the more and less bilingual subgroups was on the modified Bogardus scale: the more bilingual group showed reliably less own group preference than did those with little facility in English, $t_{52} = 2.43$.

DISCUSSION

Evaluational reactions. It is not unexpected that English speaking *S*s should show more favorableness to members of their linguistic group, but the finding that French *S*s also evaluate English guises more favorably is as unexpected as the finding that these *S*s judge French guises less favorably than do the English *S*s.

The traits given to *S*s for evaluation included several of those commonly considered necessary for social and economic success (e.g., looks, leadership, confidence, ambition), but we attempted to balance these with personality characteristics generally considered of greater value such as likability and kindness. It is possible that we were biased in the selection of traits and, through ignorance, presented a list which omitted characteristics which have value for French speaking Canadians. However, the French sample chose dependability, intelligence, and kindness as the most desirable traits for friends of those given them for consideration, and two of these were rated more favorably with the English guises, a fact which argues against the view that no opportunity was given the French sample to evaluate the speakers on culturally important traits.

In view of the greater probability of finding

English people in more powerful social and economic positions in the Montreal community, both samples of *Ss* might more likely think of an English speaker as having higher status and then evaluate the English voice of the putative lawyer as having more ambition, intelligence, etc., than the French voice of the putative store employee. If the French *Ss* had a greater tendency to do so, this might account in part for their evaluational reaction pattern. Data were available to test this notion in an indirect fashion by making use of *Ss*' estimations of each speaker's likely occupation in his French and English guises. These occupational assignments were categorized as either professional or nonprofessional by two judges working independently. For the English sample, the ascribed status of English guises was markedly higher than for French guises, $\chi^2 = 38.14$ with 1 *df*. For the French sample, the same trend was noted but was much reduced in magnitude, $\chi^2 = 5.38$. The fact that the French *Ss* ascribe higher status to a larger proportion of French guises (44%) than do English *Ss* (33%) and also evaluate the French guises less favorably than do English *Ss* argues against an interpretation in terms of differences in perceived status of the two groups.

The findings can also be interpreted as meaning that the French *Ss* actually perceive members of the English speaking group as having various desirable personality traits. The characteristics ascribed to the English speakers are generally the same as those ascribed by the English *Ss* to the same guises with the exception of the English-French conflict over which group is more kind. The French *Ss* may well regard themselves as members of an inferior group, one that is nonetheless kind and religious. Several recent studies (Adelson, 1953; Sarnoff, 1951; Steckler, 1957) which document the manner in which minority groups sometimes adopt the stereotyped values of majority groups indicate that a parallel process may be taking place in the French Canadian community (cf. Stewart & Blackburn, 1956).

An examination of the common descriptions given by the French *Ss* in the open-ended questionnaire suggests that they have ambivalent feelings about both French Canadians and English Canadians, an interpretation that is analogous to the attitude of American Jews toward both Jews and non-Jews as described by Lewin (1948, p. 145ff). The French *Ss* describe their own group as progressing and having a potential (18 occurrences); feeling inferior or without confidence (16); having good qualities and being likable (16); being uninformed or uneducated (16); being intelligent or as intelligent as English Canadians (14); being individualistic or disunited (11); lacking ambition (11); and as being prejudiced or narrow-minded (10). They describe the English Canadians as being likable, good, or nice (27); thinking badly of French Canadians (19); feeling superior to French Canadians (13); being too Americanized (12); lacking knowledge of or appreciation for French Canadians (11); and as being less mannered or more rude (11).

The findings presented here take on somewhat more significance in view of two pilot studies using other samples of *Ss* which show essentially the same results. In the first study, 17 English speaking college students listened to three of the four speakers used in the present study. In this case, English guises were evaluated more favorably on all five traits presented. In the second study, 22 male and 25 female French speaking members of Catholic religious orders were *Ss* and three of the four speakers and 12 of the 14 scales used in the present study were administered. *Ss* evaluated the English guises significantly more favorably on most traits except for kindness (*bonté*) on which French guises were more favorably evaluated. Speaker Tri, as in the present study, was not subject to this trend.

Attitude data. The general trend of correlations indicates that certain measures of attitude, particularly attitudes toward one's own group, are associated with the favorableness of evaluational reactions to the spoken languages, but this relation is clearly not a marked one, especially for the English *Ss*. Furthermore, the prediction was not supported that more skill with the other group's language, which would permit more intimate interaction, would lead to smaller differences in favorableness of evaluational responses to the guises.

In view of the striking tendency for both English and French speaking *Ss* to evaluate the English guises more favorably, a possible

reason for the generally low correlations may be that generalized characteristics of French speaking and English speaking people are so widely accepted in the Montreal community that even those English *S*s with positive attitudes toward French may still perceive them as inferior on many traits. Likewise, French *S*s with negative attitudes toward English speaking people may still perceive them as superior in many respects. That is, both French and English *S*s regardless of their attitude may have come to believe that English speaking Canadians are taller and possibly more "intelligent" in the sense that they have more educational opportunities. To this extent, the attitude factor would be submerged by powerful community-wide stereotypes. This argument is limited, however, in that attitudes would certainly be expected to play the dominant role when the two guises are compared on such traits as good looks or character. To check this possible interpretation, correlations for the English sample were computed between the comparative favorableness of reactions to the guises and attitudes toward English Canadians for *each* trait examined separately. The correlations for character, intelligence, and dependability reach significance at about the .05 level but none of the others do, even those for good looks or kindness which would also be expected to reflect an attitude factor. We conclude, therefore, that the comparatively unfavorable perception of French speakers is essentially independent of the perceivers' attitudes toward French and English groups. The correlations for the French sample are not large or consistent enough to make French *S*s an exception to this generalization. Furthermore, experience with both groups, as inferred from the degree of skill in using both languages, also appears to be independent of the tendency to down grade French speaking individuals. Further research is necessary to determine the influence of community-wide stereotypes and majority-minority feelings of members of the two language groups in Montreal.

Although the findings may be considered consistent and reliable, the generality of the results will become clear only after extensions to other samples of *S*s and a larger sample of speakers. It will be recalled that the four speakers used here were chosen to represent different linguistic communities: three speakers were educated in French Canadian universities but Leo spoke more naturally with a caricatured French Canadian accent. Only speaker Tri spoke French with a Parisian accent. A comparison of the evaluational reactions made to these speakers suggests that further study of accent differences of speakers from different linguistic communities may well extend the usefulness of the technique developed here. It can be seen in Table 1 that speaker Tri was not downgraded by the French *S*s when speaking French as were the other speakers; in fact he was considered significantly more self-confident in French than in English. Nor do the English *S*s treat him adversely in French, for they perceive him as having reliably more humor, entertainingness, and sociability in French than English. Clearly more than one speaker with a Parisian speech style is needed to test this notion, but the data suggest that the Montreal community may differentiate between Parisian and Canadian French in their attitudes. Note also in Table 1 that speaker Leo, the caricatured French Canadian, is particularly downgraded in his French guise by both English and French *S*s, if we interpret the size of *t* values as an index of differential favorableness of evaluation.

SUMMARY

Samples of French speaking and English speaking Montreal students were asked to evaluate the personality characteristics of 10 speakers, some speaking in French, others in English. They were not made aware that actually bilinguals were used as speakers so that the evaluational reactions to the two language guises could be matched for each speaker. The comparative favorableness of evaluations of the two guises was correlated with various measures of attitudes toward French and English Canadians for both samples of *S*s.

The results indicate that English *S*s evaluate the English guises more favorably on most traits. French *S*s not only evaluate the English guises more favorably than French guises, but their evaluations of French guises are reliably less favorable than those of English *S*s. This

finding is interpreted as evidence for a minority group reaction on the part of the French sample.

The correlations between comparative favorableness of English and French guises and measures of attitude toward own and other groups were generally low and insignificant, especially for the English sample. The essential independence of evaluational reactions to spoken languages and attitudes is interpreted as a reflection of the influence of community-wide stereotypes of English and French speaking Canadians.

REFERENCES

ADELSON, J. B. A study of minority group authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 477–485.

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

GARDNER, R. C., & LAMBERT, W. E. Social-motivational aspects of second-language acquisition. 1958. (Mimeo.)

LEWIN, K. *Resolving social conflicts*. New York: Harper, 1948.

LICKLIDER, J. C. R., & MILLER, G. A. The perception of speech. In S. S. Stevens (Ed.) *Handbook of experimental psychology*. New York: Wiley, 1951.

SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-semitism among Jews. *J. Pers.*, 1951, **20**, 199–218.

STECKLER, G. A. Authoritarian ideology in Negro college students. *J. abnorm. soc. Psychol.*, 1957, **54**, 396–399.

STEWART, J., & BLACKBURN, J. Tensions between English-speaking and French-speaking Canadians. *Contributions à l'Etude des Sciences de l'Homme*, 1956, **3**, 145–167.

(Received October 20, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

ACHIEVEMENT MOTIVE AND TEST ANXIETY CONCEIVED AS MOTIVE TO APPROACH SUCCESS AND MOTIVE TO AVOID FAILURE

JOHN W. ATKINSON[1] AND GEORGE H. LITWIN[2]
University of Michigan

SINCE 1950 there have been an increasing number of studies using thematic apperceptive (or equivalent) measures of n Achievement (McClelland, Atkinson, Clark, & Lowell, 1953; Atkinson, 1958; French, 1958) and the Mandler-Sarason Test Anxiety Questionnaire (TAQ) (Mandler & Sarason, 1952; Sarason, Mandler & Craighill, 1952; Mandler & Cowen, 1958), which amply demonstrate that knowledge of motivational differences enhances prediction of achievement related performances. The reader who is familiar with this literature will have noted that studies of n Achievement show it to be a motive that generally enhances performance in achievement situations while studies of Test Anxiety show it to be a motive that normally produces decrements in achievement test performance. Interpretations of the results of studies that have employed only a measure of n Achievement have often included some reference to *Ss* having low n Achievement scores as apparently more fearful of failure (Atkinson, 1953; Brown, 1953; Moulton, Raphelson, Kristofferson, & Atkinson, 1958). On the other hand, studies that have employed only the measure of Test Anxiety (Mandler & Sarason, 1952; Sarason & Mandler, 1952; Sarason, et al., 1952) have viewed the behavior of *Ss* having low Test Anxiety scores as less conflicted and more task oriented. One might be tempted to think that these two sets of research findings overlap completely and that the measures employed merely tap opposite ends of a single motivational variable. This would seem a plausible inference in light of the correlation of −.43 between the two measures reported by Raphelson (1957) in a study that also showed each to be related in the expected direction to a physiological index of manifest anxiety in a stressful test situation.

The position taken here, however, is that the measure of n Achievement obtained from thematic apperception (McClelland, et al., 1953; Atkinson, 1958) or the French Test of Insight (French, 1955, 1958; French & Thomas, 1958), a very similar projective instrument, and the measure of Test Anxiety obtained from the self-knowledge scale developed by Mandler and Sarason are not measures of the same variable. Rather, it is assumed that these particular measures of n Achievement indicate the strength of a motive to approach success, while Mandler-Sarason Test Anxiety scores indicate the strength of a motive to avoid failure.

A theoretical model explaining how these two motives influence risk taking (i.e., selection) and level of performance in achievement test situations has been presented by Atkinson (1957, 1958) elsewhere. It provides a basis for testing assumptions about what is being measured by the two instruments. According to the theory, when a person's motive to achieve success is stronger than his motive to avoid failure, the resultant of the conflict is always positive; i.e., approach motivation, no matter what the level of difficulty of the task. An individual so motivated is most attracted to tasks of intermediate difficulty where the subjective probability of success is .50. Here the resultant positive motivation is strongest. However, if the motive to avoid failure (which is presumed to be a disposition to become anxious about failure under achievement stress) is stronger, then the resultant of the approach-avoidance conflict is avoidant motivation for all levels of difficulty. The maximum strength of avoidant motivation, i.e., the strongest anxiety about failure, occurs when tasks are of intermediate difficulty. Hence, according to the theory, a person so motivated finds all achievement tasks unattractive, particularly ones of intermediate difficulty.

[1] This research was supported in part by a grant from the Ford Foundation for Studies of Personality Dynamics for which the authors are very appreciative. A preliminary report of this study was given at the APA meeting in Washington in 1958.

[2] Now at Harvard University.

He performs them only when constrained by social pressures (otherwise he would "leave the field"),[3] and he performs them inefficiently; i.e., he suffers a performance decrement as a consequence of conflict engendered by competing avoidant tendencies, to the extent that his anxiety is aroused (Sarason, et al., 1952).

The present study tests several hypotheses derived from the theoretical model and simultaneously, the assumptions that (*a*) n Achievement, as measured by thematic apperception or a similar projective instrument, is a positive disposition to approach success and that (*b*) what has been called Test Anxiety is a disposition to avoid failure. The dependent variables are: goal setting (level of aspiration) when individuals are confronted with tasks which differ in difficulty (ring toss from distances ranging from 1 to 15 ft.); persistence at an achievement task when it is not clear that success has been attained (length of time spent working on a final examination); and efficiency of performance, or level of accomplishment (score on the final examination).

The hypotheses are as follows: Persons in whom the motive to achieve success is stronger than the motive to avoid failure (*a*) more often select tasks of intermediate difficulty, (*b*) work for a longer time on the final examination, and (*c*) get higher scores on the final examination than persons in whom the motive to avoid failure is the stronger motive. Hence, given the previously stated assumptions about the meaning of measures derived from the two tests, it is expected that n Achievement scores obtained from the French Test of Insight should be positively related to preference for intermediate risk, persistence, and efficiency of performance, and that scores from the TAQ should be negatively related to these same variables. Predictions should be enhanced when *S*s are simultaneously classified as high or low on both measures.

A secondary interest is to explore the the behavioral correlates of another measure called n Achievement which is obtained from the Edwards Personal Preference Schedule (1954). To date, there have been strong indications that direct (i.e., assent or preference) and indirect (i.e., imaginative or projective) methods of measuring achievement motivation do not yield comparable results. DeCharms, Morrison, Reitman & McClelland (1955) have found this to be true when a direct and indirect method of measuring n Achievement were compared in terms of predictions to behavior. Birney (see McClelland, 1958a) has found that measures of n Achievement obtained from thematic apperception and the Personal Preference Schedule (PPS) are uncorrelated ($r = -.002$, $N = 300$). This absence of relationship between the two measures is also reported by Marlowe (1959) who found, in addition, that the thematic apperceptive measure of n Achievement was positively related to peer group ratings of achievement related behavior, while the measure obtained from the PPS was not.

In summary, the purpose of this investigation is to examine the construct validity (see Cronbach & Meehl, 1955) of three contemporary measures of achievement related motives. Predictions concerning the behavioral correlates of the measures employed are generated by a theory that conceives the achievement motive and the motive to avoid failure as independent, latent, directional dispositions and states the conditions *under* which and the degree *to* which they are aroused and manifested in overt action.

METHOD

Subjects

The *S*s were male students enrolled in a sophomore-junior level psychology course at the University of Michigan in the fall of 1957. The data were collected in the laboratory sections of the course. The instructors served as *E*s.[4] There were 49 *S*s on whom both the measures of n Achievement and Test Anxiety were available. Not all of these *S*s, however, appeared for all of the later data collection sessions. In all comparisons, the largest possible number of *S*s is used, even though this results in slight fluctuations in the size of samples among the various comparisons.

Measure of n Achievement and Test Anxiety

Early in the semester *S*s were administered, under neutral classroom conditions, a Test of Insight developed by Elizabeth French (1958). The test consists of 10 short verbal statements about people which *S*s

[3] "Social pressures" reduce to other kinds of motivation, extrinsic to achievement, for performing the tasks (e.g., desire for social approval), which are not normally measured in experiments of this sort.

[4] We wish to acknowledge the assistance of Joseph Veroff and Lawrence Littig.

are asked to analyze. Both Forms I and II of the Test of Insight were used, half of the *Ss* getting each form. The interpretations of behavior, which are very similar to TAT stories, were scored for n Achievement (according to the method of content analysis described by McClelland, et al. 1953; Atkinson, 1958) by a scorer whose reliability had been established above .90 on training materials.[5] The distribution of n Achievement scores for each form was broken at the median. *Ss* above the median were classified as high in n Achievement and those below as low in n Achievement.

Test Anxiety scores were obtained from the Mandler-Sarason TAQ (see Mandler & Cowen, 1958). The TAQ were distributed in class and *Ss* were asked to complete them outside of class and to return them the following week. The questionnaires were scored by dividing each scale into intervals, assigning a score of 1–4 for each scale, and summing the scale scores to obtain the Test Anxiety score. This scoring method correlates very highly with that originally described by Mandler and Sarason ($r = .94$, $N = 50$), and differs from that recently described by Mandler and Cowen only in that it uses 4 intervals instead of 10. The distribution of Test Anxiety scores was dichotomized at the median. The PPS developed by Edwards (1954) was administered to *Ss* in class. The distribution of PPS n Achievement scores was also dichotomized at the median.

Goal Setting in Modified Ring Toss Game

Several weeks after the individual difference measures had been obtained, *Ss* played a ring toss game. Four games were set up in the room when they entered, one on each side, one in the front, and one in the back of the room. Each game consisted of a wooden peg 2 in. in diameter and 12½ in. high mounted on a round wooden base, a ring 10 in. in diameter, and 15 lines marked on the floor at one ft. intervals. A number was marked on the floor next to each line. The closest line was one ft. from the target and numbered 1; the farthest line was 15 ft. from the target and numbered 15.

The *Ss* were told: "Today you are going to play a ring toss game. You will have an opportunity to take 10 shots at the target from any line you wish. You may move after each shot or shoot from the same line. Someone will record your shots and get your code number when you finish. We want to see how good you are at this." The instructor then kept track of the shots and hits of the first few *Ss*, and these *Ss* then kept score for the rest of the class. Two of the ring toss games were used by men and two by women. There was a good deal of informal banter during the session. No attempt was made to interfere with or to control the lifelike situation. Only the results for men are considered here since there remain unresolved questions concerning the validity of measures of n Achievement and Test Anxiety when applied to women.

Data from Final Examination

The final examination in the course from which the *Ss* were drawn was a multiple choice and short answer test held several months later. It was scheduled to last from 9:00 to 12:00 a.m. Most of the students did not take the full three hours. As each student left the examination room after turning in his exam, and when he was out of sight of students still working on the exam, he was handed a card with the time written on it and asked to write his code number on it. The code number made it possible to relate the amount of time spent working on the exam to the individual difference measures, to goal setting behavior in the ring toss game, and to the final examination scores. The data were transcribed into a score for each *S* which represented the number of minutes spent working on the final examination.[6]

RESULTS

Descriptive analysis of goal-setting data

The *Ss* were classified simultaneously as high or low n Achievement and Test Anxiety. It follows from our assumptions about the two measures that the group classified high n Achievement-low Test Anxiety should be more strongly motivated to approach success than to avoid failure. The low n Achievement-high Test Anxiety group, on the other hand, should be more strongly motivated to avoid failure than any other group. The other two groups, the high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group should be more conflicted or intermediate in resultant motivation.

Figure 1 presents a smoothed curve for each of the four groups describing the percentage of shots taken from each distance. The curves are smoothed by the method of running averages. From Fig. 1 it can be seen that all four groups show some preference for the distances around 9-11 feet. However, *Ss* high in n Achievement and low in Test Anxiety prefer this region most strongly, while the low n Achievement–high Text Anxiety group shows the weakest preference for it. The other two groups are between these two as predicted. At both the closest and farthest distances, the low n Achievement Test Anxiety group takes more shots than the high n Achievement-low Text Anxiety group. The curves indicate a relative difference

[5] We wish to acknowledge the assistance of Lois Hendrickson who scored the protocols. Because many of the interpretations had no achievement imagery, the Unrelated Imagery category (UI) was scored 0 instead of −1.

[6] Because of overflow in the main examination room, 20 to 30 students took the examination in a different room. A constant of 5 min. was subtracted from the score of these students since they began the examination about 5 min. later.

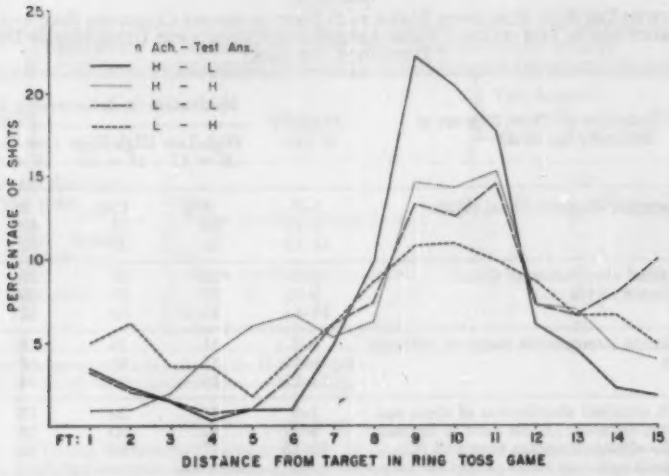


FIG. 1. PERCENTAGE OF SHOTS TAKEN FROM EACH LINE. GRAPH IS SMOOTHED ACCORDING TO THE METHOD OF RUNNING AVERAGES, FOR *S*s CLASSIFIED AS HIGH OR LOW SIMULTANEOUSLY IN N ACHIEVEMENT AND TEST ANXIETY, H-L ($N = 13$), H-H ($N = 10$), L-L ($N = 9$), L-H ($N = 13$)

between the motivation groups in preference for intermediate difficulty.

In Table 1, various ways of analyzing the data into easy, intermediate, and difficult regions are explored. The first method (I) is to divide the range of distances into equal geographic thirds. This is the method employed by Atkinson, et al. (1959) in an earlier study. The second method (II) is to use the obtained distribution of shots to determine the easy, intermediate, and difficult regions. Use of the obtained distribution of shots to define degrees of difficulty was suggested by McClelland (1958b) in an investigation of risk taking in children. The obtained distribution was first broken into approximate thirds (II-A) and then into the approximate interquartile range and extreme quarters (II-B) to show the effect of widening the region of intermediate difficulty. The third method (III) takes the middle third of distance around the obtained median of 9.8 feet.

Examination of Table 1 shows that all of the methods yield comparable results. The high n Achievement-low Test Anxiety group consistently takes the greatest percentage of shots from the intermediate region and the low n Achievement-high Test Anxiety group consistently takes the least number of shots from the intermediate region The difference. between the percentage of shots taken by these two groups from the intermediate distances ranges from 27% to 34%. The high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group fall between the two extreme groups as expected.

Effects of n Achievement and Test Anxiety

It is desirable to express the goal setting tendency of each person as a single score in statistical tests. An Average Deviation score, which was found by Litwin (1958) to be the most consistent measure across tasks of the tendency of persons to prefer intermediate risk, was used. An Average Deviation score was computed for each shot as follows: Distance of shot-Mdn. distance of all shots/ Average Deviation of all shots. The Average Deviation score for each *S* was the mean of his scores (irrespective of sign) on the 10 shots. A low Average Deviation score means that shots were taken from intermediate distances; a large score means that shots were taken from very close or very far back in relation to the median of all shots.

In Table 2 the performance of *S*s high and low in n Achievement is summarized for the three dependent variables: ring toss Average

TABLE 1

PERCENTAGE OF TEN RING-TOSS SHOTS TAKEN BY *Ss* SIMULTANEOUSLY CLASSIFIED HIGH AND LOW IN n ACHIEVEMENT AND IN TEST ANXIETY USING ALTERNATIVE CRITERIA FOR DEFINITION OF DEGREE OF DIFFICULTY (OR RISK)

Basis for Definition of Three Degrees of Difficulty (or Risk)	Distance in Feet	Motivation (n Achievement Test Anxiety)			
		High-Low $N = 13$	High-High $N = 10$	Low-Low $N = 9$	Low-High $N = 13$
I. Using geographic distance: equal thirds	1–5	8%	15%	9%	21%
	6–10	70	55	49	43
	11–15	22	30	42	36
II. Using obtained distribution of shots: A. approximate thirds	1–8	22	31	28	39
	9–10	55	39	30	25
	11–15	22	30	42	36
B. approximate interquartile range vs. extreme quarters	1–7	11	26	18	32
	8–11	73	48	48	41
	12–15	16	26	34	28
III. Using both obtained distribution of shots and geographic distance: middle third of distance about the obtained median shot (9.8 ft.)	1–7	11	26	18	32
	8–12	82	60	58	48
	13–15	7	14	24	20

TABLE 2

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *Ss* CLASSIFIED HIGH OR LOW IN n ACHIEVEMENT AND PERCENTAGE OF *Ss* ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	n Achievement		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	*N*	23	22	
	Mdn.	.63	.91	$U = 177.5$
	% below combined group Mdn.	61	36	$p = .04$
Persistence: minutes spent working on final exam.	*N*	25	19	
	Mdn.	164.9	154.0	$U = 156.5$
	% above combined group Mdn.	60	32	$p = .03$
Efficiency: final exam score	*N*	25	19	
	Mdn.	101	85	$U = 152.5$
	% above combined group Mdn.	64	32	$p = .02$

Note.—The slight fluctuation in *N* is explained in the Method section.

[a] Mann-Whitney *U* Test. All tests of significance are one-tailed tests.

Deviation score; minutes spent working on the final exam; and the score on the final exam. From Table 2 it can be seen that the high n Achievement group tends to have lower Average Deviation scores, i.e., to show a greater preference for intermediate risk, to spend more time working on the final exam, and to have higher final examination scores than the low n Achievement group. The differences between the two groups were tested by the Mann-Whitney *U* test. Since the direction of differences was predicted in each case, one-tailed tests were made. The differences between the high and low n Achievement groups are all statistically significant. Table 2 also shows the percentage of *Ss* falling above or below the combined group median in each case; 60-64% of the high n Achievement group fall in the predicted cell, and 64-68% of the low n Achievement group fall in the predicted cell.

Table 3 presents the same information for the high and low Test Anxiety groups. Again, all the results are as predicted. The high Test Anxiety group tends to have higher Average Deviation scores, i.e., to show greater avoidance of intermediate risk, to leave the final exam sooner, and to have lower final examination scores than the low Test Anxiety group. Mann-Whitney *U* tests were computed on the differences between the motivation groups;

TABLE 3

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *S*s CLASSIFED HIGH OR LOW IN TEST ANXIETY AND PERCENTAGE OF *S*s ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	Test Anxiety		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	*N*	23	22	
	Mdn.	1.00	.63	$U = 177$
	% below combined group Mdn.	35	64	$p = .04$
Persistence: minutes spent working on final exam	*N*	22	22	
	Mdn.	154.8	165.2	$U = 176$
	% above combined group Mdn.	32	64	$p = .06$
Efficiency: final exam score	*N*	22	22	
	Mdn.	91.0	101.5	$U = 177$
	% above combined group Mdn.	41	59	$p = .06$

Note.—The slight fluctuation in *N* is explained in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are one-tailed tests.

TABLE 4

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *S*s SIMULTANEOUSLY CLASSIFIED HIGH OR LOW IN n ACHIEVEMENT AND TEST ANXIETY AND PERCENTAGE OF *S*s ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	n Achievement Test Anxiety				High-Low vs. Low-High[a]
		High-Low	High-High	Low-Low	Low-High	
Risk taking: average deviation score in ring toss game	*N*	13	10	9	13	
	Mdn.	.48	1.03	.88	.93	$U = 42.5$
	% below combined group Mdn.	77	40	44	31	$p < .025$
Persistence: minutes spent working on final exam	*N*	15	10	7	12	
	Mdn.	165.3	157.5	148.3	154.5	$U = 40.5$
	% above combined group Mdn.	73	40	43	25	$p < .01$
Efficiency: final exam score	*N*	15	10	7	12	
	Mdn.	102.3	96.5	91	82.5	$U = 47.5$
	% above combined group Mdn.	67	60	43	25	$p < .025$

Note.—The slight fluctuation in *N* is explained in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are one-tailed tests.

one-tailed tests were made. The groups differ significantly in Average Deviation scores ($p = .04$). For both the number of minutes spent working on the final exam and final examination scores the probability of the difference in the predicted direction is .06. It can be seen that 59-64% of the low Test Anxiety group, and 59-68% of the high Test Anxiety group fall in the predicted cells in relation to the combined group median.[7]

[7] When only the extreme quartiles are compared, as is common in research using the Test Anxiety Questionnaire, the percentage of groups ($N = 11$) falling into predicted cells is increased somewhat to 63–72%. The median split is used here both because the *N* is relatively small to begin with and also to call attention to the relative sensitivities of the various measures of achievement related motivation when applied to the same dependent variables.

In Table 4, *S*s are again simultaneously classified on n Achievement and Test Anxiety. The high n Achievement-low Test Anxiety group is seen to have the lowest Average Deviation score (i.e., the greatest preference for intermediate risk), to spend the most time working on the final exam, and to have the highest final examination scores. The low n Achievement-high Test Anxiety group tends to have the highest Average Deviation scores, (i.e., the weakest preference for intermediate risk), to leave the final exam the earliest, and to have the lowest final examination scores. The high n Achievement-high Test Anxiety group and the low n Achievement-low Test Anxiety group generally fall between the two extreme groups as predicted. Mann-Whitney

TABLE 5

MEDIAN SCORE ON THREE DEPENDENT VARIABLES FOR *Ss* CLASSIFIED HIGH OR LOW IN PPS n ACHIEVEMENT AND PERCENTAGE OF *Ss* ABOVE OR BELOW THE COMBINED GROUP MEDIAN

Variable	Item	PPS n Achievement		High vs. Low[a]
		High	Low	
Risk taking: average deviation score in ring toss game	*N*	22	21	
	Mdn.	.97	.50	$U = 137.5$
	% below combined group Mdn.	36	62	$p = .02$
Persistence: minutes spent working on final exam	*N*	24	20	
	Mdn.	157.5	163.0	$U = 223.5$
	% above combined group Mdn.	42	55	$p = .70$
Efficiency: final exam score	*N*	24	20	
	Mdn.	100.0	89.5	$U = 181.5$
	% above combined group Mdn.	58	40	$p = .17$

Note.—The slight fluctuation in *N* is described in the Method section.
[a] Mann-Whitney *U* Test. All tests of significance are two-tailed tests.

U tests of the differences between the high n Achievement-low Test Anxiety group and the low n Achievement-high Test Anxiety group employing one-tailed alternative hypotheses are all statistically significant. When the percentage of *Ss* in each of the motivation groups falling above or below the combined group median is computed, 67-77% of the high n Achievement-low Test Anxiety group are correctly placed, and 69-75% of the low n Achievement-high Test Anxiety group are correctly placed. Simultaneous classification on both approach and avoidance motives clearly improves the prediction.

The examinations were graded A+ (excellent) to E (failure) before this analysis was even begun. It is of interest to note that the median final examination score for the high n Achievement-low Test Anxiety group (102.3) was a B−grade on the final exam, and that 67% of this group received B− or better on the exam while only 7% received a grade of D or E. On the other hand, the median final examination score for the low n Achievement-high Test Anxiety group (82.5) was a D grade; 58% of this group received a D or an E on the final exam and only 25% received a grade of B− or better.[8]

[8] These results are viewed as consistent with the idea that highly anxious individuals suffer a performance decrement on achievement tests (Sarason, et al., 1952). It may be argued that unmeasured differences in ability and not differential anxiety are responsible for the result. Perhaps *Ss* score high on Text Anxiety because they have performed badly in the past (also as a consequence of low ability). The argument is plausible, but so also is the alternative. Accordingly, finding that *Ss* who score high on Text Anxiety also score low on some independent test of academic ability would be expected since ability is always measured under the very kind of stressful test conditions that produce performance decrements in anxious *Ss*. Experimental evidence (e.g., Sarason, et al., 1952) which shows that anxious *Ss* perform as well or better than nonanxious *Ss* under nontest conditions but more poorly under achievement test conditions makes the motivational interpretation plausible pending the discovery by someone of a method of isolating the motivational and aptitudinal influences on the usual so-called test of ability.

The Measure Called n Achievement on the PPS

Table 5 presents the results for *Ss* classified as high or low on PPS n Achievement. The only statistically significant result is for the ring toss Average Deviation scores, and here it is directly opposite to what would be expected if the test measured the strength of motive to achieve success as conceived in the risk taking theory. The high PPS n Achievement group have significantly higher Average Deviation scores, (i.e., greater avoidance of intermediate risk) than the low PPS n Achievement group ($p = .02$). Two-tailed alternative hypotheses were used here since there is little evidence in previous research to support any assumptions about what this test measures. For neither time spent working on the final examination nor final examination scores are the differences between the two PPS n Achievement groups statistically significant. The trends are inconsistent. If anything, *Ss* having high PPS n Achievement scores do not work as long on the

exam but get better grades than those having low scores.

Since the PPS n Achievement measure behaves as if it were a measure of the motive to avoid failure (at least in the risk taking game), the PPS n Achievement measure was substituted for the Test Anxiety measure in a simultaneous classification with n Achievement (French Test of Insight). The high n Achievement-low PPS n Achievement group ($N = 11$) had 82% of its *S*s below the combined group median for Average Deviation score, the low n Achievement-high PPS n Achievement group ($N = 10$) had 30% of its *S*s below the combined group median. The difference between these two extreme groups was very significant ($U = 16$, $p = .01$ in this direction), larger than the comparable difference shown in Table 4 when Test Anxiety is used to assess the tendency to avoid failure!

When *S*s were classified on all three of the measures, n Achievement, Test Anxiety, and PPS n Achievement, striking differences were noted. The high n Achievement-low Test Anxiety-low PPS n Achievement group ($N = 7$) had 100% of its *S*s below the combined group median for Average Deviation score while the low n Achievement-high Test Anxiety-high PPS n Achievement group ($N = 8$) had only 25% of its members below the combined group median. The difference between these two groups was highly significant ($p = .001$).

Intercorrelation of Variables

Spearman rank order correlations were used to ascertain the degree of intercorrelation among the independent and dependent variables. The three measures of achievement related motives were not significantly related. The correlations between the several measures were as follows: Test Anxiety and PPS n Achievement, .11 ($N = 47$); PPS n Achievement and n Achievement, −.05 ($N = 47$); Test Anxiety and n Achievement, −.15 ($N = 47$).

The three dependent behavioral variables, however, tended to be positively related to one another. The index of risk taking preference, Average Deviation score (low scores given high ranks), and the measure of persistence, time spent working on the final exam, were correlated .34 ($N = 40$, $p < .05$); number of min. spent working on the final exam and final examination score were correlated .27 ($N = 40$, $p < .10$); the Average Deviation score (low scores given high ranks) and final examination scores were correlated .19 ($N = 40$, $p < .25$). A more substantial relationship is evident between the two behavioral indices of motivation, risk taking and persistence, that are not strongly influenced by academic ability, than between either of these measures and exam performance which is strongly influenced by ability.

DISCUSSION

Construct Validity of Three Tests

These results yield support for the assumptions stated earlier regarding the nature of the variables being measured by the French Test of Insight and the Mandler-Sarason TAQ. Since Litwin (1958) reports directly comparable risk taking results when using thematic apperception to assess achievement motivation, it can be concluded that n Achievement scores obtained from thematic apperception or the Test of Insight indicate the strength of a motive to approach success. Test Anxiety scores yield results consistent with the assumption that the variable involved is a disposition to avoid failure. This assumption is not at variance with Mandler and Sarason's (1952; Sarason, et al., 1952) explanation of the performance decrement shown by anxious persons under achievement stress: the anxious person's performance suffers when task-irrelevant (avoidant) responses interfere with task-relevent (approach) responses.

The results concerning efficiency of performance, or level of accomplishment, under achievement-oriented conditions replicate earlier findings concerning n Achievement (Atkinson & Reitman, 1956; French, 1955) and Test Anxiety (Mandler & Sarason, 1952; Sarason, et. al., 1952). The positive relationship between n Achievement and persistence at a task replicates an earlier finding by French and Thomas (1958) and is further evidence of the tendency also noted in recall of interrupted tasks (Atkinson, 1953). To our knowledge, this

is the first study to relate Test Anxiety to a measure of persistence. The relationship of Test Anxiety to risk taking or goal setting behavior was first noted by Vitz (1957), and Litwin (1958) has recently replicated the present risk taking findings using several different tasks.

The results concerning the measure of n Achievement obtained from the Edwards PPS are nebulous. This is the third occasion on which no relationship has been found between this direct preference measure and an indirect or projective measure of n Achievement. The observed tendency of persons having high PPS n Achievement scores to avoid intermediate risk, as if they were in fact more strongly motivated to avoid failure than to achieve, dramatically reinforces the statement that *different methods of assessing human motives do not yield comparable results* (see McClelland, 1958a). The inconsistency and lack of statistical significance of the other results pertaining to the PPS measure of n Achievement make it very difficult to understand just what is being measured by this test. One suggestive lead is provided by some unpublished results obtained by Charles Seashore.[9] He had *S*s judge the social desirability of all the statements on the PPS on a seven point scale about three months after taking the test and found a correlation of .47 ($N = 50$) between PPS n Achievement score and judged social desirability of the achievement statements. One might conclude from this that although the paired items are equated for *average* social desirability, individual differences in scores obtained from this test can be considered measures of individual differences in what is deemed socially desirable. This conclusion is supported by a recent study by Corah, Feldman, Cohen, Gruen, Meadow, and Ringwall (1958) in which judged social desirability of items within pairs on the PPS correlated highly with the item selected. Arguing in a similar vein, McClelland (1958a) has suggested designating measures obtained by direct preference or assent methods as "values", e.g., a Achievement (see also deCharms, et al., 1955), in order to avoid the inevitable confusion that arises in trying to reconcile research findings involving variables assessed by different methods but gratuitously called by the same name.

In terms of the theoretical conception of risk taking and achievement related behavior presented in detail elsewhere (Atkinson: 1957, 1958) and the assumptions regarding the variables measured here, the present findings provide considerable evidence of construct validity for the n Achievement scores obtained from the French Test of Insight and for Test Anxiety scores interpreted as measures of the strength of a disposition to be anxious about, and hence motivated to avoid, failure. However, until national norms are available for each of these measures, the crude designations as "high" and "low" of groups who stand above or below the median score in a college population are at best crude estimates of relative strength.

The relationship of n Achievement and Test Anxiety

The zero order correlation between n Achievement and Test Anxiety reported here is at variance with Raphelson's (1957) earlier report of a negative correlation. Yet the results may not, in fact, be inconsistent. In the present study the projective measure of n Achievement was presented under *neutral* classroom conditions, and the results show that persons having high n Achievement scores are as likely as not also to have high Test Anxiety scores. In Raphelson's study, the n Achievement scores were obtained from thematic apperceptive stories written by *S*s under *achievement orientation* during a rest period in the middle of a rather stressful test situation. The results of Martire (1956) and Scott (1956) clearly point to the possibility that persons who are strong in *both* achievement motive and motive to avoid failure may inhibit or distort the expression of achievement related imagery when anxiety about failure has been actively aroused in them by threatening situational cues. If so, it would be expected that persons classified high n Achievement (on the basis of *neutral* thematic apperceptive or Test of Insight scores) and also high in Test Anxiety should express less achievement related imagery when anxiety about failure is aroused. This state of affairs would induce the negative correlation

[9] Personal communication, Sept., 1958.

between the measures reported by Raphelson.

Given only high (above the median) n Achievement scores, one can assume that the achievement motive is relatively strong in relation to an unmeasured motive to avoid failure; given low n Achievement scores, one can assume that the motive to achieve success is relatively weak in relation to an unmeasured motive to avoid failure. When the motive to avoid failure is also measured, using the Test Anxiety Questionnaire, and *S*s having high n Achievement scores but also having high Test Anxiety scores are segregated, we then have more reason to be confident that our inference of relatively stronger motive to achieve success in the remaining high n Achievemet-low Test Anxiety group is correct, as the present results indicate. Similarly, segregating *S*s who have low scores on both n Achievement and Test Anxiety leaves a group low in n Achievement and high in Test Anxiety in whom we have more reason to be confident that the motive to avoid failure is relatively stronger. Yet in a group of college students it is unlikely that these are all persons in whom the absolute strength of motive to avoid failure is greater than the absolute strength of achievement motive. We find that this group does not actually *avoid* intermediate risk, as would be expected if the avoidance motive were stronger than the approach motive. Instead we find only a relatively weak preference for intermediate risk, in comparison with other groups. Presumably, most persons in whom the avoidant motive is actually stronger in an absolute sense are eliminated long before they reach college. The findings of Miller (1951), Eysenck (1955), and Eysenck and Himmelweit (1946) concerning extremely high levels of aspiration set by persons diagnosed as neurasthenic and extremely low levels of aspiration among persons diagnosed as hysteric represent the pathological extremes in avoidant motivation which may appear very rarely in college groups. The present research used all available *S*s and made its cutting point in the distributions at the median. Perhaps the choice of more extreme groups of *S*s, as has been common in research with the Mandler-Sarason test, would produce such avoidant groups even in a college population. The small number of *S*s in the present investigation precluded this possibility.

The Measure of Subjective Probability of Success

An issue of fundamental importance that is treated rather inadequately here is the measurement of subjective probability of success. Strong assumptions are made concerning how the chances of success must appear to *S*s as they approach the ring toss game. It is implicitly assumed here, as in an immediately preceding study (Atkinson, et al., 1959) which made the point explicit, that the probability of success must have appeared closer to 1.00 than to .50 at 1 ft. and closer to 0 than to .50 at 15 feet. The zone of intermediate risk or intermediate difficulty, i.e., where probability of success approximates .50, must therefore fall somewhere in between.

The actual probabilities of success obtained from the shots taken by *S*s in this experiment, while not altogether relevant since it is probability of success as it appears to *S*s at the beginning of the game that is required to test hypotheses from the risk taking theory, do nevertheless provide some support for the assumptions made. At 1 ft., the actual probability of success was 1.00. At 15 ft., it was 0. And at 7 ft., the actual probability of success was .52. However, the average observed probability of success was only .23 between 8 and 12 ft. where most shots were taken by the high n Achievement-low Test Anxiety group.

Some data collected by Litwin (1958) in a follow-up study illuminate the problem somewhat. Litwin found that when *S*s were asked to estimate their chances of hitting from each line in an 18 ft. game before performing, the average reported probability of success was nearest to .50 at 11-12 feet. This was two-thirds of the distance from the easiest to the most difficult line. In the present experiment, the modal point of shots by the high n Achievement-low Test Anxiety group was 9-10 ft., two thirds of the distance from 1 ft. to 15 feet. The *S*s in Litwin's study approached the game with overly optimistic expectancies of success. In all likelihood the same thing occurred in the present experiment, though no evidence can be presented to support the claim.

SUMMARY

It is assumed that measures of n Achievement obtained from the French Test of Insight (and thematic apperception) indicate the strength of a motive to achieve success and Mandler-Sarason Test Anxiety scores indicate the strength of a motive to avoid failure. Given these assumptions and a theory of motivational determinants of risk taking behavior, the following hypotheses are investigated: persons in whom the motive to achieve success is stronger than the motive to avoid failure (*a*) should prefer tasks of intermediate difficulty, (*b*) should show greater persistence in working at an achievement related task, and (*c*) should show more efficiency, or a higher level of accomplishment, than persons in whom the motive to avoid failure is stronger than the motive to achieve success.

Results on 49 college men regarding distance of shots in a ring toss game, time spent working on a final examination, and score on the final examination support the hypotheses and hence the construct validity of the two measures of motive strength.

Relationships between these behavioral variables and the variable called n Achievement on the PPS do not support the inference that strength of motive to achieve success is being measured by this test.

The three measures of achievement related motives are not correlated. These results highlight the importance of discovering why different methods of measuring apparently the same human motive do not yield comparable results.

REFERENCES

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.* 1953, **46**, 381–390.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

ATKINSON, J. W., BASTIAN, J. R., EARL, R. W., & LITWIN, G. H. The achievement motive, goal-setting, and probability preferences. *J. abnorm. soc. Psychol.*, 1960, **60**, 27–36.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.

CORAH, N. L., FELDMAN, M. J., COHEN, IRA S., GRUEN, W., MEADOW, A., & RINGWALL, E. A. Social desirability as a variable in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1958, **22**, 70–72.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

DECHARMS, R. C., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation.* New York: Appleton-Century-Crofts, 1955.

EDWARDS, A. L. *Edwards Personal Preference Schedule.* Manual. New York: Psychological Corp., 1954.

EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.*, 1955, **101**, 28–51.

EYSENCK, H. J., & HIMMELWEIT, H. T. An experimental study of the reactions of neurotics to experiences of success and failure. *J. gen. Psychol.*, 1946, **35**, 59–75.

FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.

FRENCH, ELIZABETH G. Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958.

FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem-solving effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

LITWIN, G. H. Motives and expectancy as determinants of preference for degrees of risk. Unpublished honors thesis, Univer. of Michigan, 1958.

MCCLELLAND, D. C. Methods of measuring human motivation. In J. W. Atkinson, (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. (a)

MCCLELLAND, D. C. Risk-taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1958. (b)

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MANDLER, G., & COWEN, JUDITH E. Test anxiety questionnaires. *J. consult. Psychol.*, 1958, **22**, 228–229.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

MARLOWE, D. Relationships among direct and indirect measures of the achievement motive and overt behavior. *J. consult. Psychol.*, 1959, **23**, 329–332.

MARTIRE, J. G. Relationships between the self concept and differences in strength and generality of achievement motive. *J. Pers.*, 1956, **24**, 364–375.

MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–387.

MOULTON, R. W., RAPHELSON, A. C., KRISTOFFERSON, A. B., & ATKINSON, J. W. The achievement motive

and perceptual sensitivity under two conditions of motive arousal. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

RALPHELSON, A. C. The relationship between imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *J. abnorm. soc. Psychol.*, 1957, **54**, 13–18.

SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SCOTT, W. A. The avoidance of threatening material in imaginative behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 338–346.

VITZ, P. The relation of aspiration to need achievement, fear of failure, incentives, and expectancies. Unpublished honors thesis, Univer. of Michigan, 1957.

(Received October 21, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOME EFFECTS OF IMMEDIATE VERSUS RANDOMLY DELAYED SHOCK ON AN INSTRUMENTAL RESPONSE AND COGNITIVE PROCESSES

M. R. D'AMATO AND W. E. GUMENIK
New York University

IT IS A common assumption that, other things equal, an unpredictable aversive stimulus has greater aversive properties than a predictable one. This general assumption often takes the more specific form that an unpredictable shock is more unpleasant, produces more anxiety, etc. than a predictable shock of equal intensity. In spite of (or perhaps because of) the rather widespread acceptance of this view, relevant data on the human level are almost nonexistent (Sidman & Boren, 1957). Confirming evidence would have an importance aside from demonstrating the obvious. It might suggest a "baseline" situation in which variables relating to individual susceptibility to both types of aversive stimulation could be studied in relation to personality variables. The experimental situation described below offers promise in this connection.

METHOD

Subjects. The *S*s were 21 male college students enrolled in the introductory psychology class.

Apparatus. The apparatus consisted of a panel on which were mounted three 7-watt lights and two response levers. A centered green light served as the "starting" stimulus, indicating the beginning of a trial. There was a right and a left response lever, each 6 in. off the midline, and a right and a left white "reinforcing" light 10 in. above the appropriate lever.

Procedure. The *S* was seated in front of the panel and orally given standard instructions, the relevant portions of which follow:

> This is a study of guessing behavior. Part of the experiment involves shock, so I'll have to put these electrodes on.
>
> When this green light goes on, you are to guess which of these two white lights will go on. Indicate your guess by pressing the lever below the light which you think will go on. This is like a gambling situation in the sense that which light goes on on any trial is determined by chance, so all you can do is guess. . . .
>
> You will be shocked during the second part of the experiment, but not during the first part. Once the shocks start, one will occur on every trial for the remainder of the experiment. Any questions?

The shock, delivered through an Applegate constant current stimulator, was then administered and adjusted to a setting which *S* sound "unpleasant but not painful." The *S* was finally requested to advise *E* if the shock became "too strong or too weak" during the course of the experiment, so that it could be readjusted. Actually, unknown to *S*, the shock was raised slightly every fifth trial of the shock portion of the experiment, a maximum of 5 milliamps. being reached in about 25% of the cases.

There were 90 trials in all. Trials 1—60 did not involve shock. The sequence of events on these trials was: green (starting) light on; as soon as *S* responded, the green light extinguished and either the right or left white light went on for 1 sec., indicating the correct response. Thirty seconds later another trial started with the onset of the green light.

Trials 61–90 involved the addition of shock on every trial, but were otherwise the same as the previous trials and were contiguous with them. One lever ("I" lever) had been preassigned to immediate shock (IS): every time *S* pressed this lever he received a shock (on the left hand) immediately after the reinforcing light extinguished, irrespective of whether he was correct or incorrect on that trial. Pressing the other lever ("D" lever) always resulted in a randomly delayed shock (RDS) of the same intensity. The delays were 0, 10, and 20 sec., occurring with relative frequencies of .20, .40, and .40, respectively. The position of the I and D levers was randomly assigned to each *S*, with the restriction that for half of the *S*s the I lever was on the right.

Throughout the experiment, the right lever was correct 50% of the time for all *S*s. Thus, regardless of *S*'s strategy, his expected percentage of correct responses was 50%. This is an important aspect of the design, since it provides for a baseline from which the effects of the independent variable may be measured. It was assumed that, by the end of the first part of the experiment (Trials 1–60), most *S*s would be responding to each lever about equally often, conforming to the reinforcement pattern prevailing. Hence, if the RDS should prove more aversive than the IS, the *S*s should press the D lever less frequently than the I lever during Trials 61–90. Of course, since the "control" nonshock Trials 1–60 are available, one could conceivably use most any distribution of reinforcement, e.g., 60:40. However, to do so could desensitize the whole procedure, since the *S*s would no doubt distribute their responses during Trials 1–60 about 60:40, and the preference for the one lever so instituted might interact with the shock variable on Trials 61–90.

At the termination of the experiment *S* was read the following questions, one by one, and his answers secured:

1. (*a*) Before shock was introduced, on what percentage of the trials did the left light go on?
   (*b*) After shock was introduced?
2. (*a*) Before shock was introduced, on what per-

centage of the trials did you press the left lever?

(*b*) After shock was introduced?

3. You probably noticed that responses to one of the levers were followed by a quick shock, while responses to the other lever resulted in a shock which was somewhat delayed for a while. If you did notice, which lever was associated with the delayed shock?

4. (*a*) Were both types of shock equally unpleasant?

(*b*) Which was more unpleasant?

(*c*) Can you tell me why?

RESULTS AND DISCUSSION

One *S* pressed the same lever on all 90 trials. His results were discarded, leaving a total of 20 *S*s.

In order to assess the effect of the shock variable on *S*s' responding, the total number of responses to the delay (D) lever was calculated for each *S* for Trials 31–60 (preshock trials) and for Trials 61–90 (shock trials). The difference between these measures was obtained for each *S*, and the (correlated) *t* test was applied to the 20 scores. The mean number of responses to the D lever for Trials 31–60 and for Trials 61–90 was 15.40 and 12.60, respectively. Although the mean difference, 2.80, was moderate in size, it proved to be reliably different from zero, $p < .01$. Thus, groupwise, pressing the lever that produced randomly delayed shock (RDS) was inhibited.

In answer to Question 4 of the questionnaire, 16 *S*s found RDS more unpleasant, two *S*s found IS more unpleasant, and for the remaining two *S*s both types of shock were equally unpleasant (Table 1). Applying the sign test to these data, we find that we may reject the hypothesis of equal unpleasantness of IS and RDS at the .01 level. Thus, both phenomenologically and with respect to its effect on an instrumental response, we have verification of the greater aversiveness of the unpredictable shock for the group as a whole.

There were, however, rather wide individual differences in the inhibiting effect of RDS. While 13 *S*s showed a decline in responding to the D lever during the shock trials, six demonstrated a reverse tendency and one *S* was unaffected. However, the largest decline obtained was 81%, while the largest increase was only 23%.

An interesting finding was the marked degree of impairment of cognitive processes which occurred during the shock trials. This impairment is illustrated by a comparison of the errors associated with Questions 1*a* ("Before shock was introduced, on what percentage of the trials did the left light go on?") and 1*b* ("After shock was introduced?"). We calculated for each *S* the absolute error involved in each of his judgments and found that while the mean absolute error for Question 1*a* was 5.6, it was more than twice as large for Question 1*b*, viz., 11.8. This increase in error associated with the shock trials is not quite significant by the two-tail Wilcoxon matched-pairs signed ranks test ($p \approx .07$). (Note that three *S*s answered Question 3 incorrectly.)

Pursuing the point further, we examined the results of Questions 2*a* ("Before shock was introduced, on what percentage of the trials did you press the left lever?") and 2*b* ("After shock was introduced?"), and once again found a decline in the accuracy of judgments made for the shock trials. The mean absolute error for the preshock trials was 6.0 and rose to 14.4 for the shock trials. Applying the previous test, we find that the difference is significant at less than the .01 level.

Thus, the conclusion seems warranted that

TABLE 1

SUMMARY OF THE SUBJECTS' RESPONSES

S	*S*s' Estimate of % of Trials D Lever Was Correct		*S*s' Estimate of % of Responses to D Lever		Actual % of Responses to D Lever		Type of Shock Found More Averaive
	Preshock	Shock	Preshock	Shock	Preshock	Shock	
1[a]	65	55	45	50	48	50	RDS
2	50	10	50	10	62	43	RDS
3	50	50	50	50	42	37	RDS
4[a]	40	55	56	50	57	66	RDS
5	50	60	50	25	43	27	RDS
6	50	50	50	15	55	30	RDS
7	50	45	50	25	55	30	RDS
8	40	80	60	20	55	37	RDS
9	50	25	50	20	43	50	RDS
10	40	50	40	40	50	47	RDS
11	50	50	65	75	53	53	RDS
12[a]	65	50	60	75	53	60	RDS
13	43	20	50	10	53	37	RDS
14	50	25	50	10	53	27	RDS
15	50	50	40	40	40	30	Equal
16	50	50	50	50	57	50	IS
17	40	50	60	50	57	43	RDS
18	60	60	50	80	50	53	IS
19	35	20	30	40	47	60	Equal
20	40	70	50	5	55	10	RDS
Means	48.4	46.3	50.3	37.0	51.4	42.0	

[a] Confusion as to D lever.

something present during the shock trials disrupted the cognitive processes necessary for subsequent accurate recall. It is important to note that both recall of external stimulus events and recall of *S*'s own behavior suffer about equally from such disruption. This fact argues against a "distraction" interpretation, in the sense of failure of the stimulus to register due to faulty receptor orientation. Clearly, such an interpretation could not apply to recall of *S*'s own behavior.

Unfortunately, we are unable to determine the relative contribution of RDS and IS to the decline in accuracy of recall, a question of considerable interest. We also cannot evaluate the factor of unequal samples (60 preshock trials versus 30 shock trials). But in view of the fact that the shock trials had the benefit of recency and considering some of the enormous errors involved, it is doubtful that this factor can be of great significance.

Closer inspection of the data revealed some interesting patterns which, because of the ad hoc nature of their emergence and the small number of *S*s involved, can only be suggestive. We wish, nevertheless, to present them at this time for their heuristic value, inasmuch as we do not plan to return to this problem.

For the purposes of the following analysis, *S*'s instrumental responses and his answers to the questionnaire were converted from left and right responses to responses to the D and the I lever. Thus, for example, the figures reported in Columns 4 and 5 (Table 1) were obtained from the answers to Questions 2*a* and 2*b*. Similarly, the figures in Columns 2 and 3 are based on the answers to Questions 1*a* and 1*b*. The items of the questionnaire were formulated so as to be unambiguous and as unsuggestive to *S* as possible. In all that follows, the translation from the basic data must be kept in mind, since it represents an unassessable factor.

Pattern A is illustrated by the results of *S* 9. This *S* estimated that he pressed the D lever on 20% of the shock trials and claimed that the D lever was correct on 25% of these trials, whereas of course the D lever was correct on 50% of the shock trials. One might be inclined to ascribe this gross error in estimating how frequently the D lever was correct to nothing more than *S* "rationalizing" his behavior. However, the fact is that *S* actually pressed the D lever on 50% of the shock trials (an *increase* over his rate of pressing this lever during the preshock trials). Note that similar results obtain for *S*s 2, 13, and 14, though to a lesser extent. In fact, if we consider the five *S*s who estimated the D lever to be correct on 25% or fewer of the shock trials (note that the next lowest estimate is 45%), these *S*s (2, 9, 13, 14, and 20) turn out to markedly underestimate the frequency with which they pressed the D lever during the shock trials. Their average error (the error obtained by subtracting the estimated percentage of responses to the D lever during the shock trials—Column 5—from the actual percentage—Column 7) is plus 22.4, whereas the average error for the entire group is only plus 5. These *S*s, as a matter of fact, yield five of the six most extreme positive errors. One might speculate that the four *S*s named found the RDS much more aversive than the IS and in part resisted its effect on the instrumental response, but only at the cost of severe distortion of recall of both their behavior during the shock trials and the stimulus events thereof.

Pattern B seems to depict a more "adaptive" approach to the experimental situation. These *S*s find the RDS more aversive than the IS and protect themselves by avoiding it. Further, they recognize their avoidance and in general maintain good perception and recall of their behavior and environment. This pattern is well illustrated by *S* 7. Note the avoidance of the D lever during the shock trials and the accuracy of this *S*'s estimates. *S* 5, and perhaps to a lesser extent 6, also display this general pattern.

Pattern C seems rather prevalent, namely, *S* shows fair to good accuracy in his estimates, claims his behavior was the same in the shock and preshock trials, but actually shows a slight avoidance of the D lever during the shock trials. *S* 16 typifies this pattern, also displayed by *S*s 3, 10, and 15. A reasonable interpretation is that the RDS is only slightly more aversive to these *S*s, if at all, and its effect on the instrumental response is too slight either to be noticed or recalled. Note that only two of these four *S*s admit to the RDS being more unpleasant.

Other behavior worthy of note is *S* 18's claim that the IS was more unpleasant, his

unaffected instrumental behavior, and the abnormally high estimate of the frequency of his D lever presses during the shock trials.

In conclusion, the authors were impressed by the magnitude and variety of effects observed under the experimental variable, particularly in view of the unselected and homogeneous sample employed. We view the above experimental situation as potentially fruitful for the study of how differential susceptibility to the effects of RDS and particular patterns of behavior, as illustrated above, are related to personality dimensions.

SUMMARY

Twenty *S*s were required on 90 trials to guess whether a right or a left light would go on, indicating their guess by pressing a lever below the appropriate light. Trials 1–60 were nonshock trials. During Trials 61–90, *S* received an immediate shock (IS) every time he pressed one lever and a randomly delayed shock (RDS) every time he pressed the other lever, irrespective of whether he guessed correctly or incorrectly. Comparison of *S*s' behavior on the shock and nonshock trials revealed:

1. A significant avoidance of the lever leading to RDS;
2. A significantly greater number of *S*s judged the RDS as more unpleasant than the IS;
3. A marked impairment of *S*s' recall of stimulus events and their behavior during the shock trials;
4. Several suggestive patterns of adjustment to the experimental situation.

REFERENCES

SIDMAN, M., & BOREN, J. J. The relative aversiveness of warning signal and shock in an avoidance situation. *J. abnorm. soc. Psychol.*, 1957, **55**, 339–344.

(Received November 7, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INCIDENTAL STIMULATION:

A STUDY OF PRECONSCIOUS TRANSFORMATIONS[1]

FRED PINE

Research Center for Mental Health, New York University

THE relevance of a stimulus to an individual's ongoing adaptive functioning is likely to be a major determinant of the nature and course of its effects. The present study deals with the participation of incidental and focal stimuli in stories that persons produce immediately following stimulus exposure. Differences in such participation related specifically to the incidental or focal character of the stimuli are explored; the main concern is with qualitative transformations undergone by incidental stimuli.

In attempting to account for the transformation from latent meaning to manifest content in dreams, Freud (1900) described a kind of thinking with "rules"—condensation, displacement, symbolization—quite out of accord with the rules of formal logic and with the experience of conscious waking thought. This mode of thinking, characterized by the "primary process," emerges in normal adult life in conditions where the reality-adaptive functions of the ego, including the ego's function of critical guidance of the direction of thinking, are in abeyance—for example, in dreams or on the analytic couch. The central hypothesis of this study is that the vicissitudes of irrelevant incidental stimuli are governed more by the laws of primary process thinking while those of relevant focal stimuli are governed more by the laws of the secondary process. Stimuli that are irrelevant to reality-adaptation may escape being used in thought that is directed towards adaptive manipulation of reality (the "secondary process") and may undergo transformations having no reality-adaptive significance, reflecting the archaic forms and highly idiosyncratic drive-related content of primary process thinking.

The present study is closely related, by way of contrast, to studies of "discrimination without awareness" (see Adams, 1957); also McConnell, Cutler, & McNeil, 1958). The focus of these studies is on what may be described as the "accuracies" in behavior with respect to a stimulus presented below the threshold of awareness. In contrast, the present study deals with the "inaccuracies" in the effects of marginal stimuli (subliminal or incidental)—effects traceable to a stimulus but nonetheless quite altered from the initial stimulus. It is also closely related to studies of the effects of subliminal stimulation which indicate that unreported stimuli emerge in cognitive productions subsequent to the stimulation (Allers & Teler, 1924; Bach & Klein, 1957; Fisher & Paul, 1959; Klein, Spence, Holt, & Gourevitch, 1958; Paul & Fisher, 1958; Poetzl, 1917; Shevrin & Luborsky, 1958; Smith, Spence & Klein, 1959). An additional observation in these studies has been that what eventually emerges in subsequent dreams or images is related to but certainly not always identical with the initial stimulus. The present study is focused specifically on these qualitative alterations in the stimulus (here, an incidental stimulus) and the theoretical implications of such stimulus transformation.

A common criticism of studies of discrimination without awareness and of subliminal stimulation has been that the stimuli were not truly subliminal and that subjects were in fact responding to some partial information stemming from the stimulus (see Adams, 1957; Bricker & Chapanis, 1953; McConnell et al., 1958). This critique has at times tended to direct theoretical discussion towards the issue of whether (and how) a truly subliminal stimulus can influence behavior. Even if the hypothesis of partial cues is helpful in explaining the influence of such stimuli, however, the hypothesis contains no direct suggestion for

[1] The study is one of a series supported by a grant from NIMH (Project Code Number: M-782, Cognitive Processes Below the Level of Awareness). I am indebted to my colleague, George S. Klein, whose thinking had important influences throughout on the basic design and the theoretical orientation of the research. The staff of the Research Center for Mental Health contributed to the study by carrying out rating tasks and by reading the manuscript. I wish especially to thank Robert R. Holt for his many suggestions and Frances Lippmann for her major share in the rating procedures and the data analysis.

explaining the qualitative changes that they undergo. The use of incidental rather than subliminal stimuli in the present study, in addition to extending related research using subliminal stimuli (cf. Klein, 1958), was intended to bypass the issue of subliminal-or-not for the moment and to refocus on the issue of qualitative differences in the effects of stimuli when they are in focal awareness as compared with when they are incidental and irrelevant to reality-adaptation.

METHOD

*S*s were 24 male undergraduates seen individually in one-hour sessions. They were asked to invent two stories (the "pre" stories). Next, two stimuli were presented simultaneously: a "focal" one that *S* read and an "incidental" one, produced in an adjoining room, that was experienced by *S* as irrelevant "noise." The stimuli were verbal descriptions of a cow and of a hook; each was incidental for half of the *S*s. Following this, *S*s again were asked to invent two stories (the "post" stories). Data were analyzed by comparing influences of the "cow" and the "hook" stimuli on the poststories under the incidental and focal conditions.

The stimuli were written by *E* to suggest particular kinds of indirect associations. It was therefore possible to predict specifically at least some of the directions of change in the content of the stimuli when administered incidentally; emergence of a transformed derivative of a stimulus rather than some intact stimulus content was expected for incidental stimuli. Psychoanalytic theory and clinical description underly the thinking about the associative value of the stimuli.

The "cow" stimulus was written to convey oral-passive connotations, using the mother-infant relationship (feeding, contact, warmth) as a model. It reads as follows:

> The ordinary cow is an animal which is indispensable to mankind. Among the domesticated animals, it holds a place of special importance for us all. The dairy industry, and in turn all consumers, are dependent upon the cow as a basic source of supply. Cows thrive in the warm climates of the temperate zones, in the protective atmosphere of small farms. They are found especially on gently rolling meadowlands where serene warmth promotes a contented existence. Lands where grass is lush and plentiful are the special home of the cow. The cow itself is a warm-blooded animal. On contact, its flesh too has a moist warmth as though the inner warm-bloodedness pervades the whole being of the animal. It is a soft animal. At its fullest development, the cow has a fleshy covering which envelops the whole of its frame. The cow is warm; it is soft. Cows live in groups. They follow one another and hover close to each other. Even when spread out across a small meadow, there is a basic "oneness" among the several cows of a small herd—all living together in the lush and warm meadow which is their home.

Any of the manifest content of the stimulus appearing in the stories was considered a relatively direct stimulus derivative. Any stray content that seemed more distantly related to the stimulus, either on theoretical grounds or on the basis of clinical impression, was considered a transformed derivative. Thus, direct derivatives could include references to cows, grass, and warm climates. Transformed derivatives could include oral-passive themes such as mothering and dependent relationships.

The "hook" stimulus was written to convey phallic-aggressive connotations, using as a model the syndrome of masculinity, aggression, intrusiveness, voyeurism, and exhibitionism which psychoanalytic writers have described as characteristic of the phallic stage of development. It reads as follows:

> If your glance has ever been captured by the silvery glint of a steel framed hook, you may have sensed its enormous potential for mechanical action. The capacity of a hook to pierce the surface of huge bulks makes it the tool of choice for use when heavy, active, and effective work must be carried out. Hooks can be moved rapidly, but there is a firm sureness to their movements when they are properly used. They hiss through the air and seise firmly upon their targets. The larger hooks can be used to raise great weights off the ground, to swing them here and thereat will, to control the movements of other objects with a steady and sure grasp. In viewing the hook, one cannot avoid being aware of its metallic hardness, its steel grey coldness, and of the stark prong at the end of the molded curve which forms the hook. The hard steel is an essential part of its makeup. The hook's unbending arc is a bulwark of the power required for the immense tasks it must perform. But the effectiveness of the cold steel is ultimately determined by the sharpness of the polished edge. Coldness, hardness, silvery steel: these are the qualities of the hook.

Direct derivatives of the hook stimulus could include references to sharpness, mechanical work, swift motion, and the like. Transformed derivatives could include, for example, aggressive, voyeuristic, or exhibitionistic themes.

The procedure, step-by-step, was as follows. *S* and *E* were alone in an office that was sufficiently isolated to virtually exclude extraneous noise. *S* had been told only that the study had to do with "imagination and some other things." To begin with, *S* was given some material from the MAPS Test (Shneidman, 1951). A card with a picture of an empty theater stage was used along with four blank-faced figures usually seen as boy, girl, man, and woman. *S*'s task was to build a scene on the stage using any or all of the four figures and then to tell a story about it. *S* was then given the blank TAT card and was asked to tell another story. Standard TAT instructions were given for both stories; instructions were written out and given to *S* to eliminate any need for a potentially biasing inquiry. All stories were tape recorded and later transcribed.

Following the storytelling, *S* was told that he would get a rest from imagination tasks while *E* got a measure of his "ordinary reading rate" in reading "just for comprehension." Instructions were for *S* to read a passage silently several times, announcing each time a reading

was completed so that *E* could note the time. *S* was encouraged to read at a steady rate even as he got increasingly familiar with the material after the first reading; he was told that the aim of the study was to examine "consistency" in reading rate. This instruction succeeded in keeping *S*s well focused on the reading task even though the incidental stimulus was audible from an adjoining office simultaneously. *S* was given either the cow or the hook passage. *E* had *S* read this over and over, at least until the broadcast of the incidental stimulus was completed; 18 *S*s read the focal stimulus just three times and the remaining 6 *S*s read it two to five times.[3] After the reading, *E* reviewed the salient points in the passage for *S* in a standardized manner under the guise of making certain *S* clearly understood them; the intention was to maximize the focal quality of the passage.

Ten seconds after *S* began reading the focal stimulus, *E* (unbeknown to *S*) signaled an assistant in an adjoining room to broadcast the incidental stimulus. The incidental stimulus (cow or hook) was the one other than the one *S* was reading. It was tape recorded by a male voice (not *E*'s) and always played at the same volume. The voice volume was such that it could be heard in the experimental room as clearly as though it were being read at close to ordinary speaking volume in the same room as *S*. The incidental stimulus took just under two minutes to read; it was read once. The subgroups of the study are hereafter referred to by the incidental stimulus they received; there were 12 *S*s each in the "cow-incidental" and in the "hook-incidental" groups.

Various devices were used to increase the likelihood that *S* would experience the incidental stimulus as irrelevant. When *E* and *S* first arrived at the experimental room, the door was always "accidentally" locked and *E* had to get the key from someone who was "working" next door. Then, during *S*'s telling of the pre-experimental stories, the phrase "testing—1, 2, 3" was broadcast several times on two separate occasions. Finally, the incidental stimulus itself was preceded by a reading of "testing—1, 2, 3" five times; it was found in discussion with pretest *S*s that this gave *S* a chance to experience the interruption to his own (focal) reading and to accustom himself to the intrusion even before the incidental stimulus itself was broadcast.

Following the reading task, *S* was told that he would now "return to the imagination experiment" and was instructed to tell stories once again. The same MAPS Test material was given, followed by the blank TAT card. Story instructions for the poststories were identical with those for the prestories except that *S* was told to tell a different story from the one he had told the first time.

Memory for the focal and incidental stimuli was next investigated. *S* was asked to tell all he could recall of the material he had read. He was then given a recognition test consisting of 42 phrases, half of which had appeared in the written passage; *S* checked off those he recognized. Finally, *S* was reminded that he "might remember that there was some interfering noise while he read the passage earlier. He was then given the same recall and recognition tests for the incidental stimulus.

Generally, *S*s were quite surprised when they were asked about the incidental stimulus. Only one *S* recalled more than one or two important words (e.g., cow, skin, metal, silvery), although five *S*s were able to recognize correctly from three to seven phrases. Only one *S* said that he thought that the incidental stimulus might have something to do with him, and no *S*s suspected any connection between the story-telling and the incidental and focal stimuli. The incidental stimulus was typically experienced as an interference, sometimes with annoyance, and attempts were made to ignore it. Memory for the incidental stimulus was thus quite poor even though it was well above the auditory threshold. More significant from the point of view of the study, however, is the fact that *S*s experienced the incidental stimulus as irrelevant.

[3] The data for these six *S*s suggests that the more the focal stimulus was read, the less was the effect of the incidental stimulus.

RESULTS

Several analyses of the stories were undertaken, each designed to test the central hypothesis that irrelevant incidental stimuli, in contrast to relevant focal stimuli, tend to undergo transformations reflecting the archaic forms and drive-related content of primary process thinking.

Quality of main relationship. Story content that could potentially reflect the emergence of transformed derivatives of either the incidental or the focal stimulus was first examined. It was expected that such story content would show the effects of the incidental rather than the focal stimulus. The quality of the relationships between the characters in the stories was analyzed in accord with this expectation.

Interpersonal relationships were rated for their oral-passive and phallic-aggressive qualities. Such drive-related qualities as these in relationships between people are related to the peripheral connotations of the cow and hook stimuli, respectively, but are only indirectly related to the manifest stimulus content; they were conceived of as transformed derivatives of the cow and hook stimuli. It was predicted, therefore, that oral-passive and phallic-aggressive qualities in interpersonal relationships would be influenced by incidental rather than focal stimuli.

The interpersonal relationships in the stories were analyzed in three parallel ways. In each, a continuum was described ranging from categories with connections primarily with the

(transformed) cow stimulus through to those with connections primarily with the (transformed) hook stimulus. The categories are given here along with the score (in parentheses) assigned to each; the numerically higher scores were always assigned to those categories conceived to be related to the cow stimulus. In the first analysis, after Erikson (1950), relationships were categorized as "receiving" (3), "giving" (2), "taking" or "withholding" (1), and "intruding" (0). In the second analysis, after Horney (1937), relationships were categorized as "moving towards" (2), "moving away from" (1), and "moving against" (0). In the third analysis, relationships were categorized as "positive" (2), "mixed" (1), and "negative" (0). These three systems of analysis are obviously not independent; for example, a "positive" relationship is likely to be "receiving" or "giving" and unlikely to be "moving against." The results give different ways of looking at one over-all finding.[3]

The placement of stories into categories was carried out by a rater who knew neither which stories were "pre" and which "post" nor which stimulus was incidental and which focal. For each of the three systems of analysis, two scores were derived for each *S*: a "post" score derived by adding up the category scores for the two post stories, and a "difference" score reflecting the change in score from the two pre- to the two poststories. The 24 *S*s were ranked on each of the three categorizations, and differences between the cow-incidental and hook-incidental subgroups were tested by the Mann-Whitney test. Results are presented in Table 1.

All results were in the expected direction. Findings were strongest for the receiving–intruding dimension where a finer categorization of relationships in the stories made for a wider range of scores. Thus, stories of *S*s in the cow-incidental group tended to involve relationships at the "receiving" end of the continuum, while stories of *S*s in the hook-incidental group tended to involve relationships at the "intruding" end. An examination of the category analyses from another viewpoint indicated further that, on all three systems of categorization, *S*s in the cow-incidental group showed an absolute increase of cow-related interpersonal relationships in the post- over the prestories, while *S*s in the hook-incidental group showed an absolute increase of hook-related interpersonal relationships.

TABLE 1

SIGNIFICANCE OF DIFFERENCES BETWEEN RANKS OF COW-INCIDENTAL AND HOOK-INCIDENTAL GROUPS ON INTERPERSONAL RELATIONSHIPS CATEGORIES

Rating Dimension	"Post" scores		"Difference" Scores	
	z Score[a]	p Value[b]	z Score[a]	p Value[b]
Receiving–Intruding	2.80	.01	1.82	.05
Towards–Against	1.01	.16	1.36	.09
Positive–Negative	1.39	.09	1.85	.05

[a] All differences are in the predicted direction.

[b] One-tailed tests.

These results are especially striking since all *S*s received both the cow and the hook stimuli. Furthermore, *S* read the focal stimulus several times in contrast to hearing the incidental stimulus only once. In spite of this, or as it will later be argued, *because* of this, the effects just described were related to the incidental stimulus and went in a direction opposite to the focal stimulus.

Cast of characters. The kinds of characters used in the stories were also examined to study the effects of the experimental stimuli. The oral-passive connotations of the cow stimulus could produce a tendency towards the use of children as characters in the subsequently produced stories. An effect of this kind is only indirectly related to the manifest content of the cow stimulus and was conceived of as a transformed stimulus derivative; as such, these effects were expected to show the influence of incidental rather than focal stimuli. It was predicted, therefore, that the oral-passive connotations of the cow stimulus would make for greater use of children in stories of *S*s in the cow-incidental group in contrast to the

[3] These abstract category titles may give a misleading impression regarding how directly or indirectly the interpersonal relationships are related to the manifest content of the cow and hook stimuli. For example, "moving towards" was typically represented in the stories not by people standing near each other (a more direct derivative) but in a wide range of idiosyncratic ways, quite unrelated to the manifest content of the cow stimulus—for example, a man buying his son a new car or a family making plans for a vacation.

greater use of adults in stories of *S*s in the hook-incidental group.

The presence of men, women, boys, and girls was tabulated for all stories; only one tabulation was made for each of these four types in each story regardless of the number of any one type (because of the frequent use of groups of people as story characters where the number in the group was unspecified). The ratio of child types (boys or girls) to the total number of types used in the two poststories, and also the difference in this ratio from pre- to poststories, was computed for all *S*s. Scores were ranked and the Mann-Whitney test (one-tailed) was used to test the significance of the differences between ranks in the two subgroups. Results were in accord with initial expectations. The use of children in the poststories was significantly greater in the cow-incidental group ($z = 2.02$; $p = .025$); the converse of this regarding the greater use of adults by *S*s in the hook-incidental group is also true. The same trend also appeared in the pre-to-post difference scores ($z = 1.44$; $p = .075$). Absolute changes were in accord with the relative rank differences; in the cow-incidental group there was an increase in the ratio of child types in the two poststories as compared to the two prestories, and in the hook-incidental group there was an increase in the use of adult types.

In a further analysis, a rater who did not know which stimulus was incidental and which focal judged whether a child or adult was used as the single principal character in each of the poststories. In this case, since no decision could be made for several of the stories, the scores for use of child or adult as the main character are not directly complementary to one another. Separate rankings were made for all *S*s of the number of poststories in which a child was used as the main character and also for the the number which used an adult. The Mann-Whitney test was again used to test for differences. More children were used as main characters by *S*s in the cow-incidental group ($z = 1.65$; $p = .05$), and more adults by *S*s in the hook-incidental group ($z = 1.53$; $p = .07$). This result supports the one reported above for ratio of children and adults used, but is clearly not independent of it. The two analyses give different ways of looking at a single finding. However, these child-adult findings are conceptually distinct from the earlier reported quality of main relationship findings; relationships involving either children or adults could be either, for example, positive, mixed, or negative. Both of these over-all sets of findings demonstrate, from one point of view, the emergence of the incidental stimuli in transformed ways and, from another point of view, the greater effect of incidental than of focal stimuli on aspects of the stories which are only indirectly related to the manifest stimulus content.

The data on the cast of characters of stories were also analyzed in terms of male and female, but no significant differences between subgroups were obtained. It had been expected that the use of male characters, a transformed derivative of the (phallic-aggressive) hook stimulus, would be more prevalent in the hook-incidental group while the use of female characters would be more prevalent in the cow-incidental group. One may speculate that in a male sample with strong tendencies to use males as identification figures in the stories, an effect of variations in immediate stimulus conditions on sex of story characters cannot readily emerge.

Over-all incidental-focal judgments. A global analysis of the story content was also undertaken. The stories were submitted to a group of raters who were asked to judge which stimulus had been incidental and which focal. Each of the 24 protocols of four stories was judged by three different raters. One rater judged all 24 protocols, while the second and third raters varied from protocol to protocol; 15 raters were used in all. Raters were instructed to make their judgments by using the hypothesis that the incidental stimulus would emerge in the stories in more transformed ways and the focal in more direct ways. Raters were informed which two stories were "pre" and which two were "post"; they were advised to use the prestories as a base line for revealing characteristics of the stories attributable to personality factors in *S* rather than to the stimulus.

Kinds of story content which raters singled out as stimulus derivatives and used in determining the final judgments may be illustrated as follows: Items conceived of as

relatively direct derivatives of the cow stimulus included phrases such as "nice countryside," "playing in a field," "at the market," "average," and "foodstuffs," as well as themes such as life on the countryside and animal stories—all of which can be traced in a careful reading of the cow stimulus. Relatively transformed derivatives of the cow stimulus included phrases like "holding his stomach," "gaining sympathy," and "at the dining room," as well as themes such as a mother protecting her son, and receiving welfare benefits from the state. The rationale for the selection of transformed derivatives varied with the judgment of the individual rater; in the examples just given, the oral-passive quality of the derivatives was a determining factor. Relatively direct derivatives of the hook stimulus included such phrases as "hurrying," "working hard," "glancing around him," and "effective work" as well as themes such as repairing things and going to engineering school. Relatively transformed derivatives of the hook stimulus included phrases such as "becomes successful" and "a cold snowstorm" as well as themes of aggression, death, voyeurism, and exhibitionism.

Whenever at least two of the three raters judged correctly which stimulus had been incidental for an *S*, the combined judgment was considered correct. (Since one judgment was missing for one *S* and the remaining two judges split, this analysis is based on an *N* of 23.) Of 23 *S*s, 17 were judged correctly ($p = .02$, Fisher exact test). Judges were able to do equally well for both the cow-incidental and the hook-incidental subgroups. Nine of 12 judged correctly in the cow-incidental group ($p = .07$) and 8 of 11 in the hook-incidental group ($p = .11$); smaller sample size here accounts for the failure to reach the .05 level.[4]

DISCUSSION

Qualitative differences thus appeared in the effects of relevant focal stimuli and of irrelevant incidental stimuli on subsequently produced stories. Incidental stimuli tended to emerge in transformed ways in the stories. Were one to look only for intact, rather than transformed, derivatives of incidental stimuli, some of the major effects of such stimuli would be obscured. A recent study (Goldstein, 1958) in which direct measures of the registration of subliminal stimuli proved less "sensitive" than indirect measures supports the same view. It was suggested, regarding the central hypothesis of the present study, that results could be explained by postulating that different streams of thought intervene between stimulus exposure and the emergence of effects in conditions of incidental and focal stimulation. Psychoanalytic theory (cf. Freud, 1900; also Rapaport, 1951) gives the fullest description of such contrasting thought streams, subsumed under the concepts of "primary process" and "secondary process."

The focal stimuli of the present study fall under the sway of adaptive intentions. An *S* instructed to read for comprehension functions adaptively by trying to grasp the literal content of the cow or the hook passage. When recall of the focal stimulus is demanded, adaptation and communication are only served by reporting on the cow as a cow, on the hook as a hook—and nothing else. It is precisely this direction of thinking by adaptive intentions, including the intentions to communicate effectively and to cope adequately with the requirements of external reality, which defines the secondary process. It may be suggested further that once a focal stimulus has achieved a clear-cut definition in full awareness it does not as readily reappear in distorted ways in imaginal productions; such a process would parallel the common experience wherein, once structure has been given to a vague stimulus, it can no longer be seen readily in any other way.

In contrast to the focal stimuli, the incidental stimuli were irrelevant to the *S*s' tasks; they appear to have entered only minimally into awareness and adaptive thinking. It is unlikely, in this setting, that the cow and hook stimuli attained a fixed and literal

[4] Two additional analyses of the data were carried out using some further judgments made by the raters who gave the incidental-focal judgments. These involved a listing of all recognizable derivatives of either the cow or the hook passages in the stories, and ratings of the degree of orality and of aggression of the stories. Scores derived here proved to be unreliable. However, statistical analysis was carried out with these scores and did not confirm the hypothesis that incidental stimuli emerge in more transformed ways.

identity comparable to the clarity and definiteness of concepts in secondary process; the transformed quality of the derivatives reflect a fluctuating and many-sided definition of the incidental stimuli. By using as a model both the drive-directedness and the nonrational forms of primary process thinking, we were able to predict some of the specific effects of the incidental stimuli. It may be that the peripheral oral-passive and phallic-aggressive connotations of the stimuli activate drives in the person that tend towards discharge by a "spilling over" of drive-related content into the stories. In addition, it may be that the incidental stimulus has the quality of an incompleted task (cf. Zeigarnik, 1927) which, never having become fully structured in awareness, undergoes a continuing and relatively autonomous process of transformation within the person (cf. Rapaport, 1957). Thus, what emerges in the story protocols as transformed derivatives of incidental stimuli may actually be the end result of some associative process. The incidental stimuli set off a chain of association to other material relevant to specific drives which have been activated and along pathways with such nonrational formal links as negation, clang association, and symbolization.

What are the conditions for a stimulus to lead to such associations and thus to emerge only in transformed ways in later thinking? At least two are suggested.

First, when a stimulus is irrelevant for adaptive functioning the likelihood of its entering into the meandering stream of primary process thought is increased. Incidentalness of a stimulus may at times be associated with irrelevance, for example as the incidental stimuli were in the present study, but it is by no means true that all incidental stimuli are irrelevant for adaptive functioning. Whether decreased awareness and decreased relevance produce effects similar in all respects to one another is unlikely but that at least some of their effects are similar is suggested by the correspondence of the results of the present study with those of studies of subliminal stimulation. The subliminal stimulus (see Fisher & Paul, 1959; Shevrin & Luborsky, 1958) does not generally tend to produce direct representations of itself (as advertisers would like to have it), but instead emerges in quite transformed ways.

The use of imaginative story production, and probably of fantasy in general, as a vehicle to carry the emerging derivatives of the stimulus also heightens the likelihood of stimulus transformation. Fantasy, involving some relaxation in customary standards of reality orientation, throws open the doors to those thought contents which are ordinarily kept out of awareness because of their irrelevance and also permits breakdowns in logic that would not be acceptable in reality-adaptive thinking. It is probably because of the use of TAT type stories to express the stimulus derivatives that even the focal stimuli emerged with some transformation in the poststories. In contrast, memory for the focal stimuli showed almost no transformation of the stimulus content.

Even though this study employed incidental stimuli and imaginative productions, both of which facilitate emergence of derivatives suggestive of nonrational thought processes, it should be noted that the effects of the incidental stimuli were still limited. The derivatives of the incidental stimuli did not completely "capture" the themes of the subsequent fantasies. Nor did the presence of incidental stimulation markedly lessen the effectiveness of memory for focal stimuli. The special status of an irrelevant incidental stimulus gives it an advantage for certain kinds of effect but not a power transcending the reality-oriented and adaptive functioning of the ego based upon intake of relevant stimuli in full awareness and acting in lines guided by conscious intentions.

SUMMARY

The present study of effects of incidental stimulation on subsequently produced stories, which contrasted the quality of effects produced by incidental as opposed to focal stimulation, is closely related to recent studies on the effects of subliminal stimulation.

Twenty-four male *S*s were seen individually for a study of "imagination." Each began by telling two stories, with standard TAT instructions, to ambiguous stimuli. Following this, two stimuli were administered simultaneously. The "focal" stimulus was a written passage

that *S* was asked to read as a measure of his "ordinary reading speed." The "incidental" stimulus was a similar passage read by a male voice and broadcast through the wall from an adjoining room; this stimulus was consistently experienced as irrelevant "noise" by *S*s. The two passages, describing a cow and a hook, were respectively devised to convey oral-passive and phallic-aggressive connotations; each was incidental half of the time (while the other was simultaneously focal). Finally, *S* again told two stories.

Contrasting effects of the incidental and focal stimuli were sought in the poststories, using prestories as a control. The prediction that irrelevant incidental stimuli tend to emerge in the stories in more transformed ways, and relevant focal stimuli in more direct ways, was tested; this prediction was derived from the theoretical hypothesis that irrelevant stimuli enter into primary process modes of thinking more than do focal stimuli. The results gave general support to the prediction. For example, blind raters were able to judge which stimulus had been incidental and which focal by basing judgments on the prediction that incidental stimuli would tend to emerge in transformed ways.

The relation of the study to recent work on subliminal stimulation, the relation of the results to psychoanalytic theory of primary and secondary process, and the conditions for transformation of incidental stimuli were discussed.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

ALLERS, R., & TELER, J. Uber die Verwertung unbemerkter Eindrucke bei Associationen. *Z. ges. Neurol. Psychiat.*, 1924, **89**, 492–513.

BACH, S., & KLEIN, G. S. The effects of prolonged subliminal exposures of words. *Amer. Psychologist*, 1957, **12**, 397–398.

BRICKER, P. D., & CHAPANIS, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.

ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.

FISHER, C., & PAUL, I. H. The effect of subliminal stimulation on images and dreams: A validation study. *J. Amer. Psychoanal. Ass.*, 1959, **7**, 35–83.

FREUD, S. *The interpretation of dreams*. (Original German ed., 1900. New York: Basic Books, 1955.)

GOLDSTEIN, M. Subliminal stimulation: Its cue value in fantasy. Paper read at American Psychological Association, Washington, D. C., September 1958.

HORNEY, KAREN. *The neurotic personality of our time*. New York: Norton, 1937.

KLEIN, G. S. On subliminal activation. Paper read at American Psychological Association, Washington, D. C., September 1958.

KLEIN, G. S., SPENCE, D. P., HOLT, R. R., & GOUREVITCH, SUSANNA. Cognition without awareness: Subliminal influences upon conscious thought. *J. abnorm. soc. Psychol.*, 1958, **57**, 255–266.

MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

PAUL, I. H., & FISHER, C. The effects of subliminal visual stimulation on images and dreams. Paper read at American Psychological Association, Washington, D. C., September 1958.

POETZL, O. Experimentell erregte Traumbilder in ihren Beziehungen zum indirekten Sehen. *Z. ges. Neurol. Psychiat.*, 1917, **37**, 278–349.

RAPAPORT, D. Towards a theory of thinking. In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.

RAPAPORT, D. Cognitive structures. In J. S. Bruner, E. Brunswik, L. Festinger, F. Heider, K. F. Muenzinger, C. E. Osgood, & D. Rapaport, *Contemporary approaches to cognition*. Cambridge: Harvard Univer. Press, 1957.

SHEVRIN, J., & LUBORSKY, L. The measurement of preconscious perception in dreams and images: An investigation of the Poetzl phenomenon. *J. abnorm. soc. Psychol.*, 1958, **56**, 285–294.

SHNEIDMAN, E. S. *Thematic test analysis*. New York: Grune & Stratton, 1951.

SMITH, G., SPENCE, D. P., & KLEIN, G. S. Subliminal effects of verbal stimuli. *J. abnorm. soc. Psychol.*, 1959, **59**, 167–176.

ZEIGARNIK, B. Uber das Behalten von erledigten und unerledigten Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85.

(Received November 14, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

PARENTAL IDENTIFICATION, ACCEPTANCE OF AUTHORITY, AND WITHIN-SEX DIFFERENCES IN COGNITIVE BEHAVIOR[1]

JAMES BIERI

The New York School of Social Work, Columbia University

Sex differences in behavior continue to be reported in the literature of personality research, a recent example being the studies of anxiety reviewed by Jensen (1958). In the area of cognitive behavior, a rather consistent finding has been that males do better than females on certain perceptual tasks such as the embedded figures test (EFT) (Witkin, Lewis, Hertzman, Machover, Meissner, & Wapner, 1954; Bieri, Bradburn, & Galinsky, 1958). One way to conceptualize the basis of these sex differences is to analyze behavioral differences *within* each sex in terms of personality variables which, while common to both sexes, may be considered to be closely related to the sex role of the individual. Further, these personality variables should be logically related to the dispositions considered to be crucial in producing differential performance on the perceptual task being studied.

Witkin et al. (1954) consider differences in EFT performance to be due to field dependent-field independent orientations of the individual. Such a view, in turn, may readily be related to sex-role differences; that is, males are more field independent and females more field dependent. Presumably, in our culture, the male sex role involves more active, assertive, independent behaviors than does the female sex role. In addition to Witkin's own work, research by others at least partially supports this view. Thus, Marlowe (1958) has recently reported a significant negative relationship between succorance (dependency) and EFT performance.

These considerations suggest the value of analyzing within-sex differences in EFT performance as a function of two personality variables which conceptually should be related to field dependence-independence. The first of these is parental identification. It is assumed children learn behaviors more similar to those of the same sex parent. Whether this identification is achieved primarily through defensive strategies, as postulated in psychoanalytic theory, or whether it represents a more positive emulation of parental behavior, is an issue not directly of concern here. What is crucial is the assumption that a person's pattern of parental identifications is closely associated with his sex role behavior, and that these identifications should thus prove to be related to sex differences in cognitive behavior.

Assume that an individual develops identifications with both parents, but a stronger identification may exist with one parent than with the other. Assume further that characteristic sex role behaviors are associated with paternal and maternal identifications. Specifically, persons who identify with the father, in contrast to mother identifiers, should manifest greater independence, assertive and active coping behaviors, a syndrome resembling the field-independent person described by Witkin et al., (1954). In comparison, *S*s who identify primarily with the mother are assumed to manifest more field-dependent behaviors. These differential behaviors should apply to both sexes, e.g., women who identify with the father should be more field independent than women who identify with the mother.

The second personality variable hypothesized to be related to cognitive behavior is acceptance of authority (AA), one aspect of the broader authoritarian syndrome, akin to authoritarian submission as discussed by the California group (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). If AA reflects a more passive, conforming, and accepting orientation toward authority, such an orientation should be reflected in poorer or field-dependent behavior on the EFT.

These formulations lead to the following hypotheses:

[1] This study was supported by a research grant (M-1976) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

The writer wishes to acknowledge the assistance of Robin Lobeck in analyzing the data for this study.

1. Within each sex, and in both sexes combined, *S*s who identify more with the father score lower on the EFT than mother identifiers.

2. Within each sex, and in both sexes combined, *S*s who are low in AA score lower on the EFT than *S*s who are high in AA.

3. Within each sex, and in both sexes combined, *S*s who have low AA scores and who identify with the father score lower on the EFT than *S*s with high AA scores who identify with the mother.

METHOD

The *S*s consisted of 30 college undergraduates of each sex attending a large Eastern private university. Each *S* was administered the following tests in an individual session.

Embedded Figures Test. An eight-item form of Witkin's EFT was administered in standard fashion to each *S*. Previous research has demonstrated the equivalence of this form to a longer version of the EFT (Bieri et al., 1958). Since the EFT is scored for latency of response, a low score indicates better performance.

Parental identification. In a previous paper (Bieri, Lobeck & Galinsky, 1959) two aspects of the identification process were analyzed, namely, perceived similarity to a parent and degree of involvement with a parent. It was found that the measures of identification in terms of perceived similarity (interview and semantic differential) were more consistently interrelated for both sexes than were the involvement measures (interview and TAT). For this reason, the perceived similarity aspect of identification was focused upon in the current research, and the semantic differential was selected as the measuring instrument. The form used consisted of 12 bipolar adjectives that were rated on a seven-point scale. Each *S* first rated himself, then his mother, and then his father. Discrepancy (*D*) scores were calculated on each of the 12 scales, reflecting the disparity between *S*'s rating of himself and his rating of each parent. It was considered that *S* identified most strongly with that parent who had the lower *D* score.

Acceptance of authority. This variable was measured by a scale containing 20 items with high loadings on a factor called acceptance of authority by Bales & Couch (1956). The items have a mean loading of .65 on the AA factor, with a range from .36 to .76. In a previous report, where this scale is presented in detail, it was demonstrated that this scale correlates highly with the F scale (Bieri & Lobeck, 1959).

RESULTS

Contrary to expectation, the mean performance scores of the males and females on the EFT were not significantly different, the values being 32.31 and 33.03, respectively. In an earlier study (Bieri et al., 1958), using the same form of the EFT on samples of 50 males and 60 females, mean EFT scores of 27.90 and 41.04 respectively for each sex were obtained. In comparison to these earlier samples, the males in the present research thus appear to be more field dependent and, the females more field independent. Features of the present *S*s' parental identification to be discussed may have operated to produce this similarity between the sexes on the EFT.

TABLE 1

MEAN EFT SCORES OF *S*s GROUPED BY PARENTAL IDENTIFICATION AND ACCEPTANCE OF AUTHORITY

	Male	*p*	Female	*p*	Combined	*p*
Mother Ident.	36.00 (N = 13)	—	40.09 (N = 11)	.04	37.88 (N = 24)	.09
Father Ident.	27.47 (N = 15)		29.78 (N = 18)		28.73 (N = 33)	
High AA	42.00 (N = 10)	.02	37.30 (N = 10)	—	39.65 (N = 20)	.04
Low AA	21.33 (N = 9)		34.20 (N = 10)		28.10 (N = 19)	
High AA—Mother Ident.	67.25 (N = 4)	.06	44.00 (N = 4)	.10	55.62 (N = 8)	.02
Low AA—Father Ident	28.00 (N = 3)		27.00 (N = 4)		27.43 (N = 7)	

Note.—Mann-Whitney test; one-tail *p* values. One male in the low AA group inadvertently was not given the EFT.

Hypothesis 1. As predicted, females who identify with the father have significantly lower EFT scores than those who identify with the mother (Table 1). For males and both sexes combined, the difference is in the predicted direction but fails to reach the .05 significance level. It should be pointed out that contrary to expectations, more women than men identified with the father, and more men than women identified with the mother (Table 1). These characteristics of the *S*s' identifications are consistent with the interpretation that the males in this study tend to be relatively field dependent and the females tend to be field independent, and that these tendencies contributed to the minimal over-all sex differences in EFT performance that were reported above.

Hypothesis 2. Each sex group was divided into thirds, representing high, middle, and low AA scores. Hypothesis 2 was evaluated in two ways; first, by testing the significance of the difference between EFT scores of high

and low AA *Ss*, and second, by correlating AA and EFT scores for *all* *Ss* in each sex group. It should be noted that the combined-sex groups of 20 high AA *Ss* and 19 low AA *Ss* in Table 1 are almost identical to the groupings that would have resulted had all 60 *Ss* been placed in a single distribution and then divided them into high, middle and low AA groups. By this latter method, three males in the middle group would be in the high group and three females in the high group would be in the middle group. The combined-sex groups employed avoid the unequal distribution of males and females that would have resulted by this method. From Table 1 it is apparent that for both sexes *Ss* high in AA had higher EFT scores than *Ss* low in AA, although this difference is significant only in the case of males. When the sexes are combined, the difference is significant at the .04 level. Similar results for this hypothesis were obtained by the correlational analysis. For males, AA and EFT were significantly correlated ($r = .38$, $p = .05$), while for females the correlation was negligible ($r = .05$). For the combined sexes, the value of r was .28 ($p = .05$).

Hypothesis 3. This hypothesis proved difficult to test adequately because of the few *Ss* who fell into the groups needed. Previous research (Bieri & Lobeck, 1959) has indicated that *Ss* high in AA tend to identify more with the father in contrast to low AA *Ss* who identify more with the mother. Therefore, fewer *Ss* would have high AA scores and maternal identification or low AA scores and paternal identification. In spite of the few cases obtained, differences in the predicted direction were obtained (Table 1), although these differences were significant only when males and females were combined.

DISCUSSION

While the results were consistently in the predicted direction, the findings suggest that parental identification and acceptance of authority may operate differentially within each sex in relation to EFT performance. For females, parental identification seems more reliably associated with differences on the EFT than does high or low AA. This conclusion is in line with previous work with these same *Ss* (Bieri & Lobeck, 1959) which suggests that AA is a less discriminating variable in women than in men, possibly because attitudes toward authority play a more important role in the male's sex-role development.

For men, AA may be a more discriminating variable in terms of EFT performance than is parental identification. Thus, while high AA males had significantly higher EFT scores than low AA males, the most striking contrast in EFT performance was between males with high and low AA scores who identified with the mother. The mean EFT scores for these two groups were 67.25 and 18.00, respectively, a difference significant below the .01 level. In contrast, males who identify with the father but who differ in terms of having high or low AA scores, have mean EFT scores that are not reliably different (25.17 and 28.00, respectively). Thus, in both males and females, the combination of high AA and mother identification produces the most field-dependent behavior. For males, low AA and mother identification produce the most field-independent behavior, while for females, low AA and father identification produce the most field-independent behavior.

The greater field independence observed in *Ss* with cross-sex identification and low AA is perhaps consistent with the interpretations sometimes made that persons considered to have a more stable sexual identification may manifest some opposite-sex tendencies in their behavior. Thus, Witkin et al. (1954) found that field-independent *Ss* had more opposite-sex features in their drawings of same-sex humans than did field-dependent persons. Similarly, Adorno et al. (1950) consider the low authoritarian to be more secure in his sexual identification and thus able to acknowledge more opposite-sex behavior than the high authoritarian.

These findings make it desirable to revise the original assumptions of the study concerning the nature of identification in relation to cognitive behavior. The results are generally consistent with the reasoning that father identifiers are more field independent and mother identifiers are more field dependent. However, the fact that low AA males who identified with the mother had significantly lower EFT scores than high AA males who identified with the mother suggests that

maternal identification in males reflects a more passive, field-dependent orientation when combined with high acceptance of authority, but reflects a more active, field-independent orientation when combined with low acceptance of authority. Thus, not only do low AA males with maternal identification have lower EFT scores than high AA males with this identification, but in addition they are the only male group whose mean EFT score is significantly lower than the mean of the comparable female group (18.00 and 39.00, respectively; $p = .02$, two-tails).

It is apparent that due to the restricted nature of the samples of this study, and the small numbers of *Ss* involved, the present findings are best construed as providing direction for further research. In particular, it should be stressed that the study bears on only one aspect of identification, perceived similarity between oneself and one's parents. The results suggest that additional work along the lines of this study may not only yield additional information on the relationship of identification to cognitive behavior, but also clarify the nature of the identification process itself.

SUMMARY AND CONCLUSIONS

Sex differences in cognitive behavior may be conceptualized in terms of personality variables related to within-sex differences, when it is reasonable to assume these variables are related to the cognitive behavior under study. Using measures of parental identification and acceptance of authority, it was predicted that within each sex, and for both sexes combined, the following differences in performance on the embedded figures test should be found: (*a*) *Ss* who identify with the father should score lower on the EFT than mother identifiers; (*b*) *Ss* low in AA should score lower on the EFT than *Ss* who are high in AA; (*c*) *Ss* with low AA scores and who identify with the father should score lower on the EFT than *Ss* with high AA scores and who identify with the mother.

The results support Hypothesis 1 in the case of females only, although for males and for the sexes combined, the results are in the predicted direction. Hypothesis 2 is supported for males and for both sexes combined, and again the results for females are in the predicted direction. Hypothesis 3 is supported for sexes combined, although differences in the predicted direction were obtained for males and females separately.

These results suggest that parental identification may operate more simply in relation to differences in EFT performance for females, while attitudes toward acceptance of authority may operate more directly in relation to EFT performance in males. However, further analyses indicated that for males, *Ss* with low AA and maternal identification had significantly lower scores than *Ss* with high AA and maternal identification. Thus, maternal identification in males may reflect quite different cognitive orientations depending upon the level of attitudes toward authority. The relevance of this cross-sex identification in behavior was discussed in relation to these findings. Further research with these variables might add to our understanding of both the nature of personality dispositions affecting cognitive behavior and the nature of identification per se.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BALES, R. F., & COUCH, A. A factor analysis of values. Unpublished manuscript. *Lab. of Soc. Relat.*, Harvard Univer., 1956.

BIERI, J., BRADBURN, WENDY, M., & GALINSKY, M. D. Sex differences in perceptual behavior. *J. Pers.*, 1958, **26**, 1–12.

BIERI, J., LOBECK, ROBIN, & GALINSKY, M. D. A comparison of direct, indirect, and fantasy measures of identification. *J. abnorm. soc. Psychol.*, 1959, **58**, 253–258.

BIERI, J., & LOBECK, ROBIN. Acceptance of authority and parental identification. *J. Pers.*, 1959, **27**, 74–86.

JENSEN, A. R. Personality. In P. R. Farnsworth (Ed.), *Annu. Rev. Psychol.* Palo Alto, Calif.: Annual Reviews, 1958. Pp. 295–322.

MARLOWE, D. Some psychological correlates of field independence. *J. consult. Psychol.*, 1958, **22**, 334.

WITKIN, H. A., LEWIS, HELEN B., HERTZMAN, M., MACHOVER, KAREN, MEISSNER, PEARL B., & WAPNER, S. *Personality through perception.* New York: Harper, 1954.

(Received November 24, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE EFFECTS OF NEED-ORIENTED COMMUNICATIONS ON ATTITUDE CHANGE[1]

FRANCIS J. DI VESTA AND JACK C. MERWIN
Syracuse University

ONE of the processes underlying attitude change appears to consist of the internalization, or incorporation, of the ideational content of the attitude (Katz, Sarnoff, & McClintock, 1956). A theoretical framework based on this position (Woodruff & Di Vesta, 1948; Smith, 1949; Rosenberg, 1956) assumes that favorable attitudes toward an object (issues, activities, ideologies, and the like) result from perceptions (beliefs or accepted information) that the attitude object facilitates need satisfaction; and unfavorable attitudes result from perceptions that the attitude object blocks or hinders need satisfaction. Accordingly, it follows that the manipulation of need (or value) strength, of the perceived instrumental relationship of the object to the satisfaction of needs, or of both may be effective in producing a change in attitudes.

The present experiment differed from earlier investigations (Woodruff & Di Vesta, 1948; Carlson, 1956; Rosenberg, 1956; Peak & Morrison, 1958) based on this theory of attitude structure in the following ways: (*a*) change in attitude was attempted by increasing the *S*'s awareness of the instrumentality of the attitude object for attaining a specified need rather than an indirect orientation toward multiple needs or values; (*b*) both negative and positive instrumental orientations were provided, each in a separate communication; (*c*) in both communications a generally positive affective assertion toward the attitude object was maintained; (*d*) changes in instrumentality of the attitude object for satisfying the need engaged in the communication were compared to corresponding changes for needs not so engaged; (*e*) the total range of scores was involved in the analyses rather than only extreme scores; and (*f*) all measures were obtained by summated rating scales developed in accordance with the definitions emerging from the theory.

The specific hypotheses were as follows:

H1: If a communication about a given attitude object or event engages a given need, the extent to which *S*s perceive the attitude object to be instrumental for satisfying the need can be changed.

1(a) If the communication engages a need in such a way that the attitude object is viewed as favorable to the satisfaction of the need, the change in perceived instrumentality will be positive.

1(b) If the communication engages a need in such a way that the attitude object is viewed as unfavorable to the satisfaction of the need, the change in perceived instrumentality will be negative.

H2: The change in perceived instrumentality of the attitude object for satisfying the need engaged by the communication will be greater than the corresponding change for any other need.

2(a) If the need is engaged by the communication in a manner favorable to the need, there will be a greater positive change in the perceived instrumentality for that need than for any other need.

2(b) If the need is engaged by the communication in a manner unfavorable to the need, there will be a greater negative change in the perceived instrumentality for that need than for any other need.

H3: The correlation between change in attitude and change in affect induction (a

[1] This research was supported in part by the United States Air Force under Contract No. AF 41(657)-73 monitored by the Office for Social Science Programs, Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, and disposal in whole and in part by or for the United States Government. This is not an official publication under the contract. Views or opinions expressed or implied herein are not to be construed as necessarily reflecting the views or indorsement of the Department of the Air Force or of the Air Research and Development Command.

multiplicative combination of need strength and perceived instrumentality) will be greater for the need engaged by the communication than for any other need.

METHOD

Subjects

The *S*s in the main experiment were 218 freshmen enrolled in the College of Liberal Arts at Syracuse University. An additional 300 *S*s from the same population, 87 *S*s from the School of Education, and 155 high school upperclassmen were used in the various studies of reliability and validity of the scales. Eighty-six *S*s enrolled in a state teacher's college were used in a pilot study.

Procedure

The attitude object for the experiment was that of "teaching as a career." *S*s in the main experiment were volunteers who knew nothing about the purpose of the study. They were part of a larger group who had taken the perceived instrumentality scales, the attitude scale, and a needs scale, together with the testing program administered to all freshmen. Upon volunteering they were alternately assigned to one of two experimental groups or the control group. The experiment was carried out in seven sessions distributed over a three-day period. During each session there were three groups of *S*s; one group received a positive communication, one a negative communication, and one a control communication. Since three *E*s were used, the communications were rotated after each session to distribute *E* effects randomly.

At the beginning of the experimental session *S*s were informed that their help was solicited to evaluate an educational radio program, and that they would be asked for reactions to questionnaires both before and after the program was played. *S*s then responded to instruments measuring needs and attitudes. These papers were collected and the program was played. At the conclusion of the program *S*s completed the perceived instrumentality and attitude scales. A detailed description of theoretical orientation, design, and measures is available elsewhere (Di Vesta & Merwin, 1958).

Measures

Attitude scales. Our theoretical orientation distinguishes between attitude and perceived instrumentality (PI). Each of the items written for the attitude scale was screened for maximum affective and minimum cognitive content. An attitude was thus defined as a relatively stable tendency to respond with positive or negative affect to a specific referent. An example of the attitude items is: "I am sure I would enjoy teaching." *S* was instructed to tell how he personally felt about teaching as a career for himself by checking one of the six categories ranging from strongly agree to strongly disagree for each of 11 statements. Statements expressing favorable opinions and statements expressing unfavorable opinions were included.

TABLE 1
INTERCORRELATIONS OF PERCEIVED INSTRUMENTALITY AND ATTITUDE MEASURES
($N = 155$)

PI	Perceived Instrumentality			Attitude
	Affiliation	Dominance	Exhibition	
Achievement	.44	.44	.46	.24
Affiliation		.22	.39	.29
Dominance			.44	.14
Exhibition				.26

The split-half reliability coefficient (300 college freshmen) corrected by the Spearman-Brown formula was .71. The coefficient of stability of the scale was .79 for 218 freshmen over a four-month period.

Perceived instrumentality scales. The measure of this aspect of cognitive structure was obtained by the use of an instrument consisting of four separate summated rating scales. These were measures of the extent to which the *S* believed that career teaching facilitated or hindered the satisfaction of the need for achievement (PI Ach), of the need for affiliation (PI Aff), of the need for exhibition (PI Exh), and of the need for dominance (PI Dom). Each scale was composed of statements about which *S* was instructed to indicate his agreement or disagreement on a six-point scale from strongly agree to strongly disagree. An illustration of a typical item is as follows: "Teachers have more opportunity to be creative than do people in most other occupations." Items were structured to obtain the *S*'s *evaluative* behavior rather than his preferential behavior as for attitudes.

The scale concerning n Ach was used as a measure of the effectiveness of the change procedure since the communications engaged that need. The other measures of perceived instrumentality (PI Dom, PI Exh, PI Aff) were used as controls.

Initially, a 64-item instrument was developed with the items for the four scales intermixed at random. Each of the four scales included some statements indicating that career teaching had a high degree of instrumentality for satisfying the need and other statements indicating that it had a low degree of instrumentality. The instrument was administered to 155 senior high school students together with the attitude scale. The intercorrelations among the four scales and the correlation of each with the attitude scale are presented in Table 1. These coefficients indicate the relative independence of these measures.

Evidence for the validity of the scale for measuring PI Ach was obtained in the administration of the 64-item scale to 87 college students majoring in education. The hypothesis was that persons having selected teaching as a career have done so, in part, on the basis of expected satisfactions of n Ach. The mean item score on the PI Ach scale was predicted to be higher for the education majors than for the high school group. The difference in the means ($t = 4.56$) was in the predicted direction and significant ($p < .001$).

The final form of the perceived instrumentality

instrument consisted of 56 of the original 64 items. A coefficient of stability was calculated for each of the four 14-item subscales. The coefficients, based on data obtained from 73 college freshmen retested after a four-month period, were: .59 for PI Ach, .55 for PI Aff, .70 for PI Dom, and .65 for PI Exh.

Need scales. Two 40-item Likert scales for measuring the four needs were developed, patterned after statements in the *Activities Index* (Stern, 1956) and the *Edwards Personal Preference Schedule* (Edwards, 1957). One of these scales consisted of statements describing activities a person would perform in social, intellectual, and/or career situations. Using a six-point scale, *S* indicated the extent to which he would *seek* or *avoid* situations involving each activity. Each of the four subscales was composed of 10 items, and the 40 items were intermixed at random throughout the instrument. On the second instrument, 40 additional statements of activities were used. The *S* used a six-point scale to indicate, for each of these statements, the extent to which he *liked* or *disliked* the activity involved.

The equivalence of the subscales was determined by relating the corresponding scores from the two newly developed instruments. The coefficients of correlation, based on 218 college freshmen were: .62 for n Ach, .57 for n Aff, .70 for n Dom, and .68 for n Exh.

Communications

Three communications were used. One provided information that teaching had favorable consequences for satisfying the need for achievement; the second that teaching had unfavorable consequences for satisfying it; and the third was a control communication on a related issue but did not engage needs in any direct manner. In both of the experimental communications a favorable assertion toward teaching was maintained, i.e., "teaching is a good career because . . . ."

The experimental communications were prepared by the *E*s and recorded on tape by the Department of Radio and Television of Syracuse University. They were the same with respect to: (*a*) the authoritative moderator, (*b*) a panel of senior division college students, (*c*) primary sources used as the bases for comments, (*d*) the number of points covered, (*e*) the enthusiasm of speakers, and (*f*) length. All programs were introduced by an announcer of the local ratio station.

Program I. In this presentation several sources were quoted by members of the panel to provide information that teaching has positive instrumentality for satisfying n Ach. Following the statement given by each member was a short summary by the moderator. Typical statements were as follows:

> The counsellors also pointed out that teaching is one job where you can get concrete evidence of your accomplishments.
>
> . . . the teacher is continually making advancement into new frontiers.
>
> . . . teaching is an occupation involving personal accomplishment, growth, and contribution to the community.

Throughout the discussion there was continual give-and-take between the panel members (college students) and the moderator (a college professor). The running time was 14 min.

Program II. This program involved the same participants and format as Program I. The questions asked by the moderator were the same as in the first program. The orientation of all information presented was that teaching lacks instrumentality for satisfying n Ach. For example, two of the statements were:

> (*a*) (In one survey) all teachers mentioned that one of the main advantages of the teaching profession, unlike other professions, is the emphasis placed on the teacher doing his own job rather than being continually in competition with other teachers; and,
>
> (*b*) The teacher has many vacation periods during which she can rest and forget about her job—unlike the scientist who continually strives to achieve new discoveries, or the industrial engineer who may spend his vacation trying to improve a product, or even the politician who during his vacation tries to achieve a bigger following.

Again it should be noted that this communication was directed toward advantages of teaching. Like the first program, this program was aimed at "why teaching is a good career." The two programs differed only in the instrumental orientation of information presented. The running time for this program was 14 min.

Program III. A control group listened to a tape recording on the topic "Going to College." This was a question and answer program that dealt with such topics as large universities vs. small colleges, costs of going to college, obtaining college catalogs, and the like. The running time was 14 min.

Pilot Study of Programs

A pilot study of the programs was conducted with 86 freshmen of a state teacher's college. The *S*s were assigned at random to one of three groups. One of the groups listened to the positive communication and a second to the negative communication. Immediately following this activity the two experimental groups responded to the PI Ach and attitude scales. The third group responded to the same scales but did not listen to a program. No pretests were used in the pilot study.

The mean PI Ach and attitude scores for the three groups are presented in Table 2. An analysis of variance of PI Ach scores yielded an F of 10.96 ($p < .001$). The t values for the differences between means of the positive and negative groups ($t = 4.05$) and between the means of the control and negative groups ($t = 3.68$) were both significant ($p < .001$). The difference between means of the positive and control groups ($t = .80$) was not significant ($p > .05$). In view of these findings and inasmuch as the means were in the pre-

TABLE 2

MEANS OF PI ACH AND ATTITUDE SCORES ON PILOT STUDY

Experimental Treatment	*N*	PI Ach	Attitude
Positive communication	29	40.83	23.75
No communication	28	42.43	23.95
Negative communication	29	50.39	23.21

Note.—The lower the mean the more favorable the PI or attitude.

dicted order the programs were considered adequate for altering PI Ach.

An analysis of variance of attitude scores yielded an insignificant F of .75. This finding can probably be attributed to the fact that the *Ss*, having selected teaching as a career and enrolled in a teacher's college, were committed to a highly favorable attitude toward teaching.

RESULTS AND DISCUSSION

The initial homogeneity of the three groups used in the main experiment is demonstrated by the F values obtained through an analysis of variance of the scores from the pretest measures. The F value for the PI Ach measure was significant ($.05 > p > .025$), but did not cause undue complications since the obtained order of the means (positive-control-negative) was opposite to that predicted to result from the experimental procedures. A significant F value ($.05 > p > .025$) for the PI Dom measure was not attributable to differences in means alone since the chi square value obtained from the Bartlett test for homogeneity of variance was also significant ($p < .05$). The F values for the other pretest measures used in the analyses were not significant. On the basis of the evidence obtained, the groups were judged to be sufficiently homogeneous for the experiment.

TABLE 3
COMPARISON OF GROUPS ON POSTTEST MEASURES

Measure	Experimental Group					
	Positive ($N = 76$)		Control ($N = 73$)		Negative ($N = 69$)	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
Perceived Instrumentality						
Achievement	45.39	9.10	47.99	7.70	52.45	9.87
Affiliation	45.71	7.44	47.03	6.25	47.93	8.26
Dominance	41.87	7.30	43.45	8.36	44.16	7.58
Exhibition	44.12	6.40	45.25	7.23	47.32	6.81
Attitude	34.97	10.79	36.30	8.86	38.20	11.49

TABLE 4
MEANS AND t VALUES[a] OF CHANGES IN PERCEIVED INSTRUMENTALITY SCORES

Measure	Experimental Group					
	Positive		Control		Negative	
	$\bar{X}$	t	$\bar{X}$	t	$\bar{X}$	t
Perceived instrumentality						
Achievement	3.33	3.67***	0.62	0.67	−7.10	−5.87***
Affiliation	2.25	2.56*	0.10	0.12	−0.91	−0.92
Dominance	1.37	1.23	1.11	1.44	−2.73	−2.12*
Exhibition	2.01	2.81**	1.19	1.65	−3.43	−4.45***

[a] Mean changes tested against a hypothesized population mean of zero (H:$\mu = 0$).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

The means and standard deviations of scores on the posttests are presented in Table 3. The mean changes for the three groups on the PI scales are given in Table 4. A plus change in PI indicates a move toward the position advocated in the positive communication and a minus change indicates a move toward the position advocated in the negative communication. "Boomerang" effects occurred with some *Ss* in both groups.

Each of the mean changes was evaluated by a t test, against the statistical hypothesis that $\mu = 0$, in accordance with experimental hypotheses H:1(a) and H:1(b). As may be seen in Table 4 the changes for both the positive and negative groups were significant ($p < .001$). The mean change of the control group was not significant ($p > .05$). These results permit rejection of the null hypothesis and clearly support experimental hypotheses H:1, H:1(a), and H:1(b).

The mean differences between change in PI Ach and change in each of the remaining three PI scores were all in the predicted direction for the positive group. None was significant however. While these data indicate trends, they do not provide sufficient evidence to support experimental Hypothesis 2(a).

For the negative group, the mean differences between change in PI Ach and the change in each of the other three PI scores were all significant ($p < .01$) and in the predicted direction. Thus, the data support experimental Hypothesis 2(b) and permit rejection of the null hypothesis of no difference in the changes.

Further support of the general experimental hypothesis (H:2) is the lack of significant differences in the changes for the control group.

In general, the effects of the communications on the PI Ach measures were as hypothesized. Nevertheless, a generalization effect is to be noted in which the positive communication

TABLE 5
ANALYSIS OF VARIANCE OF ATTITUDE CHANGE

Source	*df*	*MS*	*F*
Between groups	2	112.73	7.05*
Within groups	215	15.99	
Total	217		

* $p < .001$.

TABLE 6
MEANS AND *t* VALUES OF CHANGES IN ATTITUDE

Experimental Group	*N*	$\bar{X}$	*t*[a]
Positive	76	1.05	2.15*
Control	73	−0.81	−1.83
Negative	69	−1.29	−2.80**

[a] $H: \mu = 0$.
* $p < .05$.
** $p < .01$.

produced a generally positive change, and the negative communication produced a generally negative change in PI.

The null hypothesis that there are no significant differences in attitude change among the three groups over the experimental period was tested by an analysis of variance summarized in Table 5. The obtained *F* value of 7.05 was significant ($p < .001$). Since the chi square from the Bartlett test of homogeneity of variance was not significant, the null hypothesis was rejected. The mean and *t* values of the attitude changes for the experimental groups are presented in Table 6. The mean changes in attitude were in the direction of the instrumental positions advocated in the negative and positive communications. Since positive affect was asserted in both communications, the change in attitude is attributed to the manipulation of perceived instrumentality. These effects are further substantiated by the fact that the mean change in the attitudes of the control group was not significant ($p > .05$). The results indicate that attitudes can be modified by a manipulation of the instrumental orientation of the attitude object for attaining or hindering the satisfaction of a single need.

A measure of the change in affect induction (AI) for each need was obtained by multiplying the need score with the corresponding PI change for each individual. For example, the score for n Ach was multiplied by the change in PI Ach; n Dom was multiplied by the change in PI Dom; and so on. In addition, a composite AI change score was obtained by an algebraic summation of the four simple AI change scores. Each AI change score was correlated with the attitude change for each treatment group and for the three groups pooled. The resulting coefficients are presented in Table 7.

TABLE 7
CORRELATIONS BETWEEN CHANGE IN ATTITUDE AND CHANGE IN AFFECT INDUCTION

Affect Induction	Experimental Group			
	Positive ($N = 76$)	Control ($N = 73$)	Negative ($N = 69$)	Pooled ($N = 218$)
Achievement	.26*	.07	.22	.26**
Affiliation	−.09	−.20	−.04	−.10
Dominance	.13	−.06	.00	.07
Exhibition	.14	−.09	.10	.11
Composite	.17	−.12	.11	.13

* $p < .05$.
** $p < .001$.

Only the correlations between change in affect induction for n Ach and attitude change for the positive and pooled groups were significant. While these results would appear to provide partial support for Hypothesis 3 the relationships are too low to be considered conclusive. While the hypothesis was not advanced in the present study, it would be consistent with the theory for the composite change in affect induction to be more highly related to attitude change than would changes in affect induction based on single needs. This expectation was not supported by the data and leads to a conclusion similar to that drawn by Rosenberg (1956, p. 371) for his study that "The data leave it an open question as to whether . . . 'perceived instrumentality' actually controls more variance in attitudinal affect than does 'value [need] importance.' "

SUMMARY

Several hypotheses concerning attitude change and derived from a cognitive theory of attitude structure were tested experimentally. The change procedure consisted of three tape recorded communications, each making positive affective assertions about teaching but

maintaining respective positions that the attitude object facilitated or hindered satisfaction of the need for achievement, or making assertions irrelevant to both teaching and n Ach. Each of the two experimental groups and a control group consisted of approximately 70 subjects. Changes in beliefs about the potentiality of the attitude object for satisfying the need for achievement corresponded to the direction of the instrumental position of the respective communications. Corresponding changes in attitude were demonstrated. An index based upon need strength and change in perceived instrumentality for the need engaged in the communication was found to be only slightly related to change in attitude for the pooled groups. Similar indexes based on other needs and over all needs measured were found not to be related to shifts in attitude.

REFERENCES

CARLSON, E. R. Attitude change and attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256–261.

DI VESTA, F. J., & MERWIN, J. C. The effect of need-oriented communications on attitude structure and change. USAF WADC tech. Rep., 1958, No. 58–93 (ASTIA Doc. No. AD-151-044).

EDWARDS, A. L. *Edwards Personal Preference Schedule.* New York: Psychological Corp., 1953.

KATZ, D., SARNOFF, I., & MCCLINTOCK, C. Ego-defense and attitude change. *Hum. Relat.*, 1956, **9**, 27–45.

PEAK, HELEN, & MORRISON, H. W. The acceptance of information into attitude structure. *J. abnorm. soc. Psychol.*, 1958, **57**, 127–135.

ROSENBERG, M. J. Cognitive structure and attitudinal affect. *J. abnorm. soc. Psychol.*, 1956, **53**, 367–372.

SMITH, M. B. Personal values as determinants of a political attitude. *J. Psychol.*, 1949, **28**, 477–486.

STERN, G. G. *Activities Index.* Syracuse: Psychological Res. Center, Syracuse Univer., 1956.

WOODRUFF, A. D., & DI VESTA, F. J. The relationship between values, concepts, and attitudes. *Educ. psychol. Measmt.*, 1948, **8**, 645–660.

(Received December 19, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INVOLVEMENT AND COMMUNICATION DISCREPANCY AS DETERMINANTS OF OPINION CONFORMITY[1]

PHILIP G. ZIMBARDO
Yale University

A MAJOR problem of theoretical and practical significance in the area of attitude change concerns the placement of a persuasive communication relative to the attitude position of the person to be influenced. The bulk of experimental evidence points to a positive relationship between attitude change and degree of change advocated (Fisher, Rubinstein, & Freeman, 1956; Fisher & Lubin, 1958; Goldberg, 1954; Harvey, Kelley, & Shapiro, 1957; Hovland & Pritzker, 1957; Hovland, Harvey, & Sherif, 1957). Thus, the more discrepant a communication is from the position of the recipient the greater the attitude change in the advocated direction. However, it has been suggested by Hovland et al. (1957) that this generalization may be limited to communications (*a*) within the subjects' latitude of acceptance, (*b*) about noninvolving issues, (*c*) by credible, respected communicators.

This view stems from an extrapolation by Hovland and Sherif of results from a psychophysical weight-lifting study (Sherif, Taub, & Hovland, 1958) to the situation of communication and attitude change. They feel that discrepant communications within the latitude of rejection act like anchors *beyond* the range of a series of stimuli and should produce contrast or negative change, while communications within the latitude of acceptance act like anchors *within* a series of stimuli and should produce assimilation or positive change.

A latitude of acceptance is comprised of the position on an attitude continuum that an individual "most accepts" plus those he also accepts. The latitude of rejection includes the attitude position to which the individual "most objects" plus those to which he also objects. According to Hovland and Sherif, then, it is the placement of a communication relative to the individual's latitudes of acceptance and rejection that determines the efficacy of the communication.

A more general theoretical framework in which to explore the motivational bases of opinion change in situations of social influence has recently been advanced in Festinger's (1957) theory of cognitive dissonance. A number of derivations can be made from the theory which predict the relationship between opinion change and the variables that Hovland and his co-workers have isolated empirically.

Dissonance theory assumes a basic tendency toward consistency of cognitions about oneself and about the environment. When two or more cognitive elements are psychologically inconsistent, dissonance is created. Dissonance is defined as psychological tension having drive characteristics. Thus, the existence of dissonance is accompanied by psychological discomfort and when dissonance arises, attempts are made to reduce it. A specific application of the theory to the problem under consideration in the present paper states that dissonance is created when an individual knows that another person—a communicator—holds an opinion contrary to his own (Festinger, 1957, p. 178). The magnitude of the dissonance created by the expression of a contrary opinion will increase with increases in (*a*) the credibility and attractiveness of the communicator, (*b*) the extent of the discrepancy in opinion, and (*c*) the importance of, or involvement with, the cognitive elements of the situation.

The strength of the pressure to reduce dissonance is a direct function of the dissonance created. In a social influence situation there are a number of potential channels of dissonance reduction, such as changing one's own opinion,

[1] This research is part of a dissertation presented to the faculty of the Graduate School of Yale University in candidacy for the degree of doctor of philosophy.

The author wishes to thank A. R. Cohen for the guidance he provided during every phase of this research.

Funds from the Yale Communication and Attitude Change Project, directed by C. I. Hovland, helped to support this study.

The author is indebted to Sister Thomas Albert, Dean of Albertus Magnus College, for her cooperation throughout the study.

changing the opinion of the communicator, making the communicator noncomparable to oneself, seeking further support for one's position, dissociating the source from the content of the communication, and distorting the meaning of the communication.

One of the major methods of reducing dissonance, then, is conformity or opinion change in the advocated direction. Thus, if the other avenues of dissonance reduction are restricted, opinion change should be a function of variables that intensify dissonance. We have seen that two such variables are opinion discrepancy and importance of, or involvement with, the cognitive elements of the situation.

While discrepancy and involvement play key roles in the theory, nevertheless they remain rather vague and insufficiently specified. For the present study, therefore, discrepancy was operationally defined in Hovland and Sherif's terms as latitudes of acceptance and rejection.

The variable of involvement, which has been used by most investigators to mean concern with a given *issue* because it is intrinsically involving and is related to the individual's needs and values, was used in a different sense in the present study. This naturalistic kind of issue-involvement often confounds a number of factors, such as initial differences in commitment, social support, information, etc.

Specification of involvement in terms of the relationship of one's opinion to the achievement of a desired goal should help to limit ambiguity of interpretation. This response-involvement may be viewed as the individual's concern with the consequences of his *response* or with the instrumental meaning of his opinion. The involvement may then be differentially imposed on a previously neutral opinion area, thereby controlling the contaminating factors usually associated with the selection of different levels of issue-involvement.

Thus, the following predictions can be generated by dissonance theory for the situation in which an individual receives a persuasve communication from a "positive" communicator (response alternatives other than opinion change having been experimentally minimized): (*a*) dissonance, and hence opinion conformity, will be greater for those individuals who are highly involved than for those whose involvement is low; (*b*) dissonance, and hence opinion conformity, will be greater when the communication is widely discrepant and is directed at an individual's latitude of rejection than when it is slightly discrepant and within his latitude of acceptance; (*c*) maximum dissonance, and consequent opinion conformity, will occur when the individual is highly involved (Hi-I) and the communication is within his latitude of rejection (Lat-Rej); (*d*) minimum dissonance, and consequent opinion conformity, will occur when the individual's involvement is low (Lo-I) and the communication is within his latitude of acceptance (Lat-Acc).

In the fourfold table generated by these two variables, there should be an intermediate amount of dissonance and consequent opinion change in the cells in which the main effects are mixed, i.e., high involvement with a latitude of acceptance communication, and low involvement with a latitude of rejection communication. The theory cannot predict which of the two will be more effective, since that depends on the relative efficacy of the experimental manipulation of involvement and discrepancy.

METHOD

In order to test these predictions a before-after design was used in which 80 unpaid college girls privately gave their opinions about the locus of blame in a case study of a juvenile delinquent. After their involvement was experimentally manipulated, they were exposed to the alleged opinion about the case of a girl friend with whom they were being tested. Finally, they gave their opinions of the case study again.

Procedure

Girls volunteered with a friend for a study on how friends diagnose and evaluate social issues. The written responses of each subject (*S*) of the two pairs tested simultaneously were kept confidential by means of short partitions.

A cooperative task designed to introduce the *Ss* to the situation and to highlight the cohesiveness of each pair was given first. This task required the *Ss* to list as many social problems as they could in 60 sec., each *S* alternately listing one problem. Regardless of how well they had done, they were told that their performance was good and was better than that of comparable *Ss* who were not friends. Following this task, no communication was permitted between the *Ss* at any time.

The *Ss* then filled out a general background questionnaire which included checks on the strength of their

friendship. At this time, *E* stressed the confidential nature of their responses. They were told that because of time limitations only one social problem was selected for them to work on in more detail. The problem of juvenile delinquency was selected and related to it were two tasks. The first task involved reading a short, fictitious case history of a juvenile delinquent—"the Johnny Rodrigues case." *S* was then required to indicate her opinion concerning the locus of blame for the youth's crime. This opinion was the one *E* attempted to change.

The second task, which involved judgments of delinquents from among a set of photographs, was included primarily to increase the credibility of the friend communicator.

Each *S* was informed that while she had judged only five of eight photos correctly and was rated "good," her friend had judged all of the photos correctly and had been rated "expert." The *S*s had been told that this task measured their perceptual and judgmental ability. By having the friend appear an expert on the photo task, the impact of her *opinion* about the case study could be maximized without resorting to additional information about the case.

Involvement Manipulation

Low (*Lo-I*). Half of the *S*s were told that although they should read the case study carefully, they should not expect too much of it because it was too short, limited, and not representative of juvenile delinquency in general. Therefore, it was impossible to learn anything from their reactions to the case. They were told that *E* had hoped to learn something about their basic values and personalities by means of their case study opinions, but results with other girls from the school indicated that this attempt had proven unsuccessful. However, in order to keep everything in the experiment in the same sequence, it was necessary that they read the case and give their opinions before judging the photos.

High (*Hi-I*). *E* told the other half of the *S*s that the case study they were to read was selected because, although short, it presented a wide range of problems implicitly or explicitly, and the way they responded to this typical, representative case would be a good indicator of their basic social values, their personalities, and their outlook on important life problems.

Both groups were told that the photo task would provide information of a different kind, about an ability rather than about their opinions. After reading the case material, the *S*s made the following ratings on a scale composed of nine statements (see below) which varied in the degree to which the blame for Johnny's crime was personal or environmental. They: (*a*) selected the one statement which was most acceptable (MA), i.e., which came closest to representing their own point of view; (*b*) selected all others they felt were also acceptable (A); (*c*) selected the one statement which was most objectionable (MO); (*d*) selected all other statements which were also objectionable (O); and, finally, (*e*) indicated all statements they were unsure of, or neither accepted nor rejected.

After the opinion rating sheets were collected, the *S*s indicated how *certain* they were of their most acceptable rating.

Opinions about Johnny

1. There is no question that the blame for Johnny's crime must be placed *entirely* upon his shoulders.
2. The *greatest part* of the blame for Johnny's crime rests on his shoulders.
3. While some of the blame for Johnny's crime is due to his background, he is still responsible for *most* of it.
4. Although it is hard to decide, it is probable that *slightly more* of the blame for Johnny's crime rests on his shoulders.
5. It is difficult to decide whether Johnny himself or his background is to blame for his crime.
6. Although it is hard to decide, it is probable that *slightly more* of the blame for Johnny's crime rests on his background.
7. While some of the blame for Johnny's crime is due to Johnny himself, his background is still responsible for *most* of it.
8. The *greatest part* of the blame for Johnny's crime rests on his background.
9. There is no question that the blame for Johnny's crime must be placed *entirely* upon his background.

Communication of Friend's Opinion

The *S*s were told that since they came to the experiment in pairs of friends they would probably like to know not only how they were doing but also how their friend was doing. Thus, each pair would receive information about how they were doing, but would not have any information about the other pair of *S*s being run simultaneously.

After receiving the false photo judgment scores, every *S* was informed (on a written form) that her friend was "absolutely certain" of her most acceptable rating of the case. Half of the *S*s were informed (on a rating sheet) that their friend's MA rating was slightly discrepant from their own MA rating and was at a position which they had previously indicated as being "acceptable" (Lat-Acc). The other half of the *S*s were informed that their friend's MA rating was widely discrepant from their own and at a position they had previously indicated as being "objectionable" (Lat-Rej). Conveying the information on a rating sheet similar to that originally filled out by each *S*, and simply marking the friend's alleged opinion MA, served to make salient the opinion discrepancy and reduced the possibility of misinterpretation of the communication.

In the latitude of acceptance condition the statement selected as the friend's MA was generally one scale position above or below *S*'s MA at a position *S* had initially marked "acceptable." An attempt was made to avoid attributing a neutral position to the friend. For the latitude of rejection condition the statement selected as the friend's MA was, on the average, three scale units above or below *S*'s MA, at a position *S* had

initially marked "objectionable". The friend's MA was generally placed on that side of the *S*'s MA which offered the most potential scale movement. The *S*s were randomly assigned to the discrepancy condition, while the involvement condition was alternated from session to session; all *S*s within each testing session were given the same involvement induction.

In order to reduce the possibility that *S*s would think that they could influence their friend to change *after* the experimental session, they were told that a follow-up questionnaire would be administered some time before the end of the term. Consequently, it was necessary for them not to talk to anyone about their reactions to any part of the study, especially not with their friend.

Postcommunication Ratings

Several minutes after having collected the false information, *E* said that the *S*s were to make a fresh evaluation of the case and of the photos. This was to be done because their first opinions and judgments were immediate, almost forced responses, and *E* wanted them to make new evaluations after they had considered the problem and materials in more detail. They were told that it did not matter whether or not they changed their opinions as long as they tried to make new, independent evaluations and did not try to duplicate what they had done merely to be consistent.

The *S*s then gave their opinions of the case study again, without rereading it. Checks were then made on *S*'s certainty of opinion, perceived discrepancy of friend's opinion, and degree of involvement in each of the tasks. The partitions were removed and *E* conducted an open-ended interview to inquire into the *S*s' reactions to the study. Finally, they were thanked for their cooperation and the true purpose and design of the study were revealed. They were asked (on their school "honor") not to discuss the study with any students who had not yet participated; and there is no reason to believe that the deceptions were not effective throughout the entire course of the study.

RESULTS

Evidence of the Effectiveness of the Experimental Conditions and Manipulations

Friendship. One of the basic assumptions of the study was that each pair of *S*s was very friendly. The friendship, along with the alleged expertness of the communicator, served to make the communicator highly positive. The data clearly confirm this assumption. The girls in each pair had known each other for an average of 30 months. In addition, they gave a mean rating of 1.17 on the friend "contact" scale (1.00 = "see and talk several times a day"); and made a mean rating of 1.38 on the friend "liking" scale (1.00 = "like very much"). These results are uniform throughout all experimental conditions. The fact that the *S*s publicly announced their friendship on the sign-up sheet on the bulletin board also supports this assumption.

Perception of opinion discrepancy. The *S*s in the Lat-Rej condition perceived the communication as being more discrepant than did the *S*s in the Lat-Acc condition, a difference significant beyond the .001 level. The differences in perceived discrepancy closely paralleled the actual, experimentally created differences in opinion.

Involvement in case study opinions. The *S*s in the Hi-I condition were significantly more concerned with their case study opinions than were the *S*s in the Lo-I condition. They also felt it was more important to make a sensitive judgment of the case (both p values $< .001$). No significant differences in degree of involvement were noted for the two latitude conditions.

The third check on the manipulation of involvement is a more sensitive one suggested by Hovland, Harvey, and Sherif (1957). Individuals highly involved in a given issue will have a raised threshold of acceptance of opinion statements about the issue, while less involved *S*s will have a lowered threshold of acceptance. The present results are in line with those of the above authors: highly involved *S*s reject a significantly greater number of statements than do less involved *S*s. In sum, these data indicate that the *S*s in the Hi-I condition were more involved with their case study opinions than were the Lo-I *S*s.

Involvement in photo judgments. Since the instructions regarding the photos were identical for all *S*s, it was expected that there would not be a significant difference between the two case study involvement conditions with regard to their photo involvement. The data support this assumption. One may conclude that the photo task was a method of raising the expertness of the communicator uniformly in all conditions without differentially affecting involvement with the case study opinions.

Certainty of opinion. Though the certainty measure was used in order to help anchor the communicator's position, it also provides a check on the *S*'s confidence in his own position. Since there were no significant differences

among groups in mean certainty either before or after the communication we cannot attribute differences in opinion change to differences in certainty.

Effects of Discrepancy and Involvement on Opinion Change

In addition to use of traditional measures of opinion change, such as the scale distance changed and the percentage of *Ss* changing in each direction, the scaling procedure used in the present study permits an analysis of opinion change into six components. Briefly, these positive opinion changes consist in (*a*) moving one's "most acceptable" rating in the advocated direction (*b*) moving one's "most objectionable" rating away from the advocated direction, (*c*) adding one (or more "acceptable" ratings in the direction of the communicator's position, (*d*) removing one (or more) "objectionable" ratings from positions near the communicator's position, (*e*) removing one (or more) "A" ratings from positions behind one's own MA rating, i.e., at a distance from the communicator's position, and, finally, (*f*) adding one (or more) "O" ratings to positions behind one's own MA.

It was arbitrarily decided to combine these components into a *composite change index* by assigning a weight of +1 to each type of positive change, −1 to each negative change and zero if no change occurred, and then summing algebraically. Thus, each individual could receive a composite change score varying from +6 (maximum positive change) to −6 (maximum negative change).

Composite change index. Figure 1 presents mean total change score as a function of position of the communication and strength of involvement. It is clear from the figure that (*a*) all groups showed positive change and (*b*) that opinion change increased both with increased discrepancy and with strength of involvement. The Hi-I, Lat-Rej group, in which the experimental effects were maximized, changed most, while the Lo-I, Lat-Acc group, in which the effects were minimized, changed least. The other two "mixed" groups changed an amount intermediate between these two extremes.

An analysis of variance reveals that the difference between the two discrepancy conditions is reliable beyond the .01 level, and the difference between the two involvement conditions approaches significance at the .10 level. The Hi-I, Lat-Rej group differs significantly from each of the other groups.[2] The interaction between the main effects does not approach statistical significance.

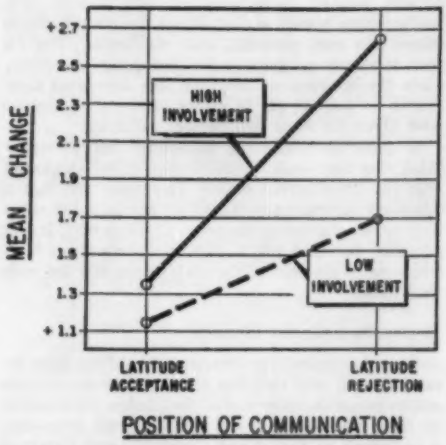


FIG. 1. Mean composite opinion change as a function of degree of involvement and extent of communication discrepancy. (Note.—Lat-Rej vs. Lat-Acc, $F = 7.74$, $p < .01$, $df = 1.76$. Hi-I vs. Lo-I, $F = 2.27$, $p = .10$, $df = 1.76$.)

Scale distance change in most acceptable rating. Table 1 presents an analysis of the data by number of scale units of net change in MA rating for each of the groups, and offers results similar to those obtained with the composite change index. There was significantly greater scale distance change in the advocated direction by *Ss* who were highly involved than by *Ss* whose involvement was low. Furthermore, significantly more change was produced by widely discrepant, latitude of rejection communications than slightly discrepant latitude of acceptance communications. Again, the relative ranking of the four cells was in accord with the predicted relationships among the groups.

Examination of the data of mean units of positive change in MA for only those *Ss* who

[2] Hi-I, Lat-Rej vs.: Lo-I, Lat-Acc, $t = 3.19$, $p < .01$; Hi-I, Lat-Acc, $t = 2.77$, $p < .01$; Lo-I, Lat-Rej, $t = 2.02$, $p < .05$, $df = 38$. All p values are for two-tailed tests of significance.

TABLE 1

MEAN UNITS OF NET CHANGE IN "MOST ACCEPTABLE" RATING

INVOLVEMENT	Position Advocated: LAT-ACC	LAT-REJ	
Hi-I	.35	.95	.65
Lo-I	.05	.30	.18
	.20	.62	

Note.—LAT-REJ vs. LAT-ACC, $F = 5.25$, $p < .05$, $df = 1.76$; Hi-I vs. Lo-I, $F = 4.19$, $p < .05$, $df = 1.76$.

actually changed their MA in a positive direction revealed trends identical with those presented for the entire groups.

Percentage of Ss changing most acceptable opinion. The results of net percentage of opinion change are in close agreement with the previous findings, e.g., the maximum dissonance group had 50% net change, while the minimum dissonance group had only 15% net change.

Equation of groups on opportunity for scale movement. Of course, the possibility must be entertained that the results for the discrepancy conditions could be explained by the methodological artifact that there was more scale distance available to change for the high discrepancy (Lat-Rej) groups than the low discrepancy (Lat-Acc) groups. Reanalysis of all the data after identically equating the groups on the degree of scale movement possible indicates that the differences obtained are accentuated rather than attenuated. For example, the difference between the discrepancy conditions on the composite change measure is significant beyond the .001 level, while the difference between the involvement condition is significant beyond the .08 level.

Relation of initial latitude size to opinion change. An unexpected finding revealed that those *Ss* who conformed initially accepted a wider range of statements and rejected a narrower range than did the nonconformers, i.e., they had larger latitudes of acceptance and smaller latitudes of rejection. This measure of "threshold of acceptance," which significantly distinguished between positive changers and those changing negatively or not at all ($p < .01$), may thus prove to be a useful indicator of susceptibility to persuasive communications.

DISCUSSION

The results of this experiment appear consistent with several general derivations that can be made from dissonance theory; principally, that under conditions specified in this experiment, opinion conformity increases as a function of involvement and discrepancy. This theory, with its consideration of the tension-producing effects of these variables, provides a sound framework for the understanding of some of the dynamics of the process of social influence.

Involvement

Although issue-involvement and response-involvement are not necessarily exclusive, they can be independently varied and should be independently conceptualized. Essentially, an experimenter is said to be using issue-involvement if the mere presentation of the issue is sufficient to elicit concern and interest. Since arousal of issue-involvement is dependent upon a cluster of antecedent factors, which generally cannot be analyzed separately, it tends to be a rather unreliable independent variable.

It seems quite plausible that under such conditions of involvement, *resistance* to change may increase as involvement increases. Hovland and Sherif[3] apparently refer to this kind of issue-involvement when they predict less positive change (and more negative change) as involvement increases. However, the results of a recent unpublished study by Brehm and Lipsher are contrary to the predictions of Hovland and Sherif and accord with the results of the present study. Brehm and Lipsher[4] manipulated involvement using issues which varied in the degree to which *Ss* were concerned with them. They found that "The

[3] Unpublished manuscript by C. I. Hovland and M. Sherif entitled "Studies in judgment."

[4] Unpublished manuscript by J. W. Brehm and D. Lipsher entitled "Communicator-communicatee discrepancy and perceived communicator trust-worthiness."

amount of *opinion change* in the direction of the communicator was greater for the issue of high concern than for the issue of low concern" (p. 11).

The other major class of involvement, response-involvement, denotes the instrumental relationship of a given opinion or response to the achievement of a desired goal. The opinion becomes a means of securing reward, approval, or recognition, of avoiding punishment, or of raising the individual's self-esteem. The involvement is not dependent upon the content of the topic per se, but is imposed upon the topic. The involvement is *created* by conditions specifying the consequences attendant upon task performance. Thus, by imposing involvement on a neutral issue, strength of involvement should be a function of strength of the experimental induction, and should be uncontaminated by differential levels of such antecedent variables as information, support, and commitment. This was a primary consideration in the present interpretation of Festinger's general statement of involvement as dealing with response involvement.

Discrepancy

The present study offers further support for the well documented finding of a positive relationship between extent of discrepancy and total amount of opinion change (Fisher et al., 1956; Fisher & Lubin, 1958; Goldberg, 1954; Harvey et al., 1957; Hovland & Pritzker, 1957). This conclusion remains valid even after controlling for regression artifacts and ceiling effects. Furthermore, its range of generalization has been broadened by the variety of ways in which it has been operationalized from discrepancies in judgments of intelligence, parachutes and dots from photos, to discrepancies in opinions on factual as well as attitudinal issues.

While replication under such diverse conditions increases the reliability of the conclusion, nevertheless the metric of change is difficult to compare across the studies because of their diversity. The measure used in the present study has a number of advantages which recommend its wider use. It is simple to administer and score, yields a description of the affective *structure* of a given opinion, and takes account of subtle changes within the structure. Thus, many of the *S*s in the present study who did not change their "most acceptable" positions showed at least one positive change in one of the components of the opinion. Therefore, measurement of changes in only the "most acceptable" position underestimates the efficacy of the persuasive communication. Similarly, one might predict vulnerability to further influence and propaganda if an individual's latitude of acceptance was extended and/or his latitude of rejection was reduced by an initial communication—even though no change occurred in his "most acceptable" position.

The sensitivity of the composite change index can be increased by use of a scaling procedure (probably Coombsian) which assigns weights on a more empirical, less arbitrary basis. The weighing system should take account of the psychological scale *distance* of each kind of change and of the relative importance of the *quality* of each component of change.

Determining the placement of communications by *S*s' latitudes of acceptance and rejection rather than by their placement at some fixed scale distance from *S*s' "most acceptable" position stems from a suggestion by Hovland and Sherif (see fn. 3). These authors feel that knowledge of the affective structure of a given opinion will aid in the optimal placement of a persuasive communication. The most obvious and rational prediction is that a person is more likely to change and accept a communication which he has previously stated he finds slightly discrepant but "acceptable," than a communication which he previously has found very discrepant and "objectionable." Indeed, the very definition of latitude of rejection implies that there are a number of positions which the individual would reject, and perhaps even respond to by changing negatively. The prediction from dissonance theory of more positive change as a consequence of communications within the latitude of rejection than of those within the latitude of acceptance is not at all obvious.

It has recently been pointed out that as discrepancy becomes extreme, a point is reached where incredulity becomes a factor mediating against opinion change. While Festinger (1957) does not explicitly specify the relationship between incredulity and discrepancy, he would

probably predict that dissonance is a monotonically increasing function of discrepancy, regardless of the extremity of the discrepancy. However, with extreme discrepancy, the prepotent dissonance-reducing response should shift from opinion change to one of dissociation, disbelief, or the discrediting of the communicator—even though experimental attempts (as in this study) are made to minimize these reactions. Therefore, dissonance theory would appear to make a prediction similar to that of Osgood and Tannenbaum (1956).

The Communicator

In the present study the communicator was highly positive for all experimental conditions; both attractive to *S* (a friend) and relevant to the issue of discrepancy (an expert on a related issue). This complex definition of positive communicator was intentionally employed in order to firmly "anchor" the position of the communicator.

While it seems obvious that "negative" communicators are generally ineffectual in producing conformity responses, under certain conditions such communicators may induce more positive change than positive communicators. Consider the situation in which an individual is "forced" to comply to a persuasive communication, i.e., to adopt a public stand or to behave in a manner contrary to his own convictions. Dissonance will be created between the sets of cognitions corresponding to his discrepant public and private positions (Festinger, 1957, Ch. 4). Subsequent (real) attitude change, as a mode of reducing the dissonance, will be an inverse function of the degree of justification for the conforming behavior. The more negative the communicator, the less the justification and the greater the dissonance produced by compliance. Thus, a negative communicator could be more influential under conditions of forced compliance than a positive communicator.

Clearly, a whole program of research is necessary in which discrepancy is varied over a wide range under various communicator conditions with several levels of both response and issue-involvement in order to answer the host of theoretical and practical questions concerning the determinants of opinion conformity.

SUMMARY

This research studied the relationship between conformity and (*a*) the extent of the discrepancy between the opinions of a communicator and a recipient and (*b*) the degree of involvement of the recipient. These variables are central to a dissonance theory analysis of the social influence process, as well as to many previous investigations of attitude change. It can be derived from the theory that opinion change increases with increases in both involvement and discrepancy between communicator and recipient.

In order to test these hypotheses, a "before-after" design was used in which 40 pairs of college girls privately gave their opinions about a case of juvenile delinquency. They were exposed to the alleged opinion of their friend (positive communicator), and then gave their opinions a second time. Response-involvement and discrepancy were experimentally manipulated in a 2 × 2 design. Attempts were made to maximize the probability of opinion change as the response to this kind of social influence.

Several different measures of opinion change (including a newly developed one) are consistent in showing that:

1. Highly involved *S*s changed significantly more than *S*s not involved.

2. Opinion change increased significantly as the extent of the discrepancy between communicator and recipient increased.

3. The interaction between these variables was not significant.

4. The maximum dissonance group changed significantly more than the other groups, while the minimum dissonance group consistently changed least. The additional finding of reliable differences between conformers and nonconformers in the relative sizes of their initial latitudes of acceptance and rejection suggests that this may be a useful index of susceptibility to change.

REFERENCES

FESTINGER, L. *A theory of cognitive dissonance.* New York: Row, Peterson, 1957.

FISHER, S., RUBINSTEIN, I., & FREEMAN, R. W. Intertrial effects of immediate self-committal in a continuous social influence situation. *J. abnorm. soc. Psychol.*, 1956, **52**, 200–207.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *J. abnorm. soc. Psychol.*, 1958, **56**, 230–238.

GOLDBERG, S. C. Three situational determinants of conformity to social norms. *J. abnorm. soc. Psychol.*, 1954, **49**, 325–329.

HARVEY, O. J., KELLEY, H. H., & SHAPIRO, M. M. Reactions to unfavorable evaluations of the self made by other persons. *J. Pers.*, 1957, **25**, 393–411.

HOVLAND, C. I., & PRITZKER, H. A. Extent of opinion change as a function of amount of change advocated. *J. abnorm. soc. Psychol.*, 1957, **54**, 257–261.

HOVLAND, C. I., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

OSGOOD, C. E., & TANNENBAUM, P. The principle of congruity and the prediction of attitude change. *Psychol. Rev.*, 1956, **62**, 42–55.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

(Received December 12, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

THE EFFECTS OF INFORMATION ON THE ASSESSMENT OF ATTITUDES BY AUTHORITARIANS AND NONAUTHORITARIANS[1]

MILTON E. LIPETZ[2]
Ohio State University

THIS study is concerned with the effects of differentially relevant information on the social judgments of authoritarians and nonauthoritarians. Previous research (Crockett & Meidinger, 1956; Scodel & Freedman, 1956; Scodel & Mussen, 1953) has indicated that authoritarians assume that others would respond like themselves to the California F scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), whereas nonauthoritarians, while more variable, tend to attribute to their partners F scale scores in the middle of the distribution from which they themselves were selected and thus tend to judge their partners as scoring higher than themselves. In these studies, judgments were made after brief interactions during which topics such as radio, television, and the movies were discussed. Crockett and Meidinger (1956) had their subjects also discuss child-rearing practices, but the amount of information made available through these interactions and its degree of relevance to the judgments remained uncontrolled. Scodel and Freedman (1956) and Christie and Cook (1958) have suggested that the inability of *S*s low in F to be more accurate may be due to the absence of detailed relevant information or to the briefness of interaction.

If *S*s high in F are governed by a stereotype of "sameness," whereas *S*s low in F are not characterized by stereotyped thinking, then the availability of information relevant to a required social judgment (in this case, judgments on the F scale) should result in greater accuracy for the latter group. The literature contains some evidence that nonauthoritarians benefit more from relevant information than do authoritarians (Juul, 1953; Kogan, 1956).

A further question was posed: If *S*s high in F but not those low in F judge others to hold attitudes similar to their own despite contrary relevant information, would measures for which the information was less relevant also reflect this differential behavior? To obtain judgments on a variable for which the information provided was less relevant than for the F scale, a set of value statements was used. Knowledge of a person's attitudes, especially relevant to judgments of his F scale score, should have intermediate relevance to judgments about his values.

It was hypothesized that, given information about relevant sample attitudes of other persons, *S*s low in F should be more accurate than *S*s high in F in differentiating a high F personality from a low F personality. Furthermore, given the same information, *S*s low in F should see their own values as more disparate from those of the high F personality than from those of the low F personality. *S*s high in F should make no such discrimination.

METHOD

Subjects

To select groups high and low in F, 28 items of the Form 40-45 California F scale were administered to 358 male college students. The scale was scored by adding a constant of four to each of the responses. Ninety-six *S*s, 48 in the high group and 48 in the low group, were selected to participate in the experiment. The high F group had total scores ranging from 121 to 148 with an item mean of 4.6. The scores of the low F group ranged from 31 to 85 with an item mean of 2.6. No restrictions were placed on the sample with respect to such factors as age, race, or religious affiliation. An estimate of intelligence was obtained for each *S* by means of the Ohio State Psychological Examination.

Measures

To obtain judgments about values, Wertheimer's six values (1957) were used. These values, each expressed in a single sentence without the value label, are Approval (A), Authoritarianism (F), Hedonism (H), Integrity (I), Other-direction (O), and Status (S). For example, the statement representing Integrity reads: One should satisfy one's own conscience, and make decisions one believes are right, regardless of whether one hurts oneself or others. Status is expressed by: One should act in such a way as to make the most of oneself, in order to earn the respect of others. Each *S*

[1] This paper is based on a doctoral dissertation submitted to Ohio State University. The writer wishes to express his sincere appreciation to his advisor, Alvin Scodel, for his encouragement and stimulating criticisms during the entire course of the dissertation.

[2] Now at the University of Colorado.

ranked the six values for himself and for the person being judged. Discrepancy scores between self values and judged values were obtained by summing the differences between the two rankings of the six values, yielding a possible range of 0–18 for any one *S*. (For example, if self values were ranked I, S, O, A, H, F, and judged values were ranked F, I, H, A, O, S, the sum of the differences is $1 + 4 + 2 + 0 + 2 + 5 = 14$.)

*S*s were also asked to fill out the F scale for the person being judged.

Source of Information

Mock interview recordings provided *S*s with information upon which to base their judgments. These recordings were made from prepared scripts whose content was derived from the theoretical clusters of Adorno et al. (1950, p. 255). One script was written to portray the high F personality and is hereafter referred to as the high role. The other represented the low F personality and is referred to as the low role. For greater representativeness four different recordings of each role were made with variations in respondents and in spontaneous and nonrelevant comments. The four persons who served as respondents in the high role recordings also served as respondents in the recordings for the low role. In each of the eight recordings the author served as the interviewer. The announced purpose of the interview was to obtain the attitudes of college students on a number of different issues as part of a research program.

In each high role recording the interviewer began by asking the respondent to express his views on juvenile delinquency. In the course of conversation the respondent expressed such opinions as: children are being pampered, children do not respect authority, and people turn to psychiatrists with their troubles instead of bearing them. In response to a question as to whether or not people should try to change the laws of the country, the respondent berated radicals and "low class" people. This interchange is designated hereafter as Information Session I. Following a brief pause the interview continued with the respondent making statements about the mysterious things going on that cannot be understood, his dislike of people who are different, his concern about sexual immorality, his belief that people cannot be changed, the need to control inferior people, and his disgust with the intermingling of races that occurred in Europe during the war. The conversation following the pause is identified as Information Session II. All the information that comprised the high role was consistent with the theoretical clusters upon which the F scale is based. An attempt was made to cover all the clusters, but in the absence of empirical data there was necessarily some subjectivity involved in determining the comprehensiveness of each role.

In each low role recording, none of the information divulged was consistent with these opinion clusters. In response to the question on juvenile delinquency the respondent replied that the problem was complicated, although the instability of society and economic factors contributed to it. At any rate severe punishment would not solve the problem. With regard to changing the laws of the country he felt that, as social conditions changed, laws had to change, although the Bill of Rights should not be diluted. During Information Session II the respondent expressed concern over social conditions, segregation, and the lack of understanding among peoples of different nations. With respect to interracial marriage he felt that this was the concern of the persons involved.

In both the low and high roles, the opinions expressed during Session II were more forceful than those offered in Session I. The adequacy of each role was informally passed on by six judges who were familiar with the theoretical clusters of the F scale. Since both roles were written to correspond to extreme positions, it is assumed that the higher a predicted score is for the high role and the lower a predicted score is for the low role, the greater the accuracy of judgment.

Procedure

From one to seven *S*s were present in any one session. They were informed that they had been selected to participate in a study investigating the value of recordings for determining the nature of social interactions. Any one *S* was exposed to only one of the eight recordings, each of which was heard by an equal number of *S*s, half high in F and half low in F. Prior to the playing of the recording, *S* was asked to rank the six Wertheimer values in order of their importance to him. The recording was then played through the end of Information Session I. Immediately thereafter, *S* was requested to fill out the F scale and to rank the six values as he thought the respondent on the recording would have answered. The recording was then continued through the end of Information Session II, after which repeated estimates of the interviewee's responses were obtained.

RESULTS

An analysis of variance of F scale judgments among records within each role by subject group and information session failed to reveal any differences significant at the .05 level. Data from the four recordings within each role were therefore combined in further computations.[3]

High Role versus Low Role

The mean estimate scores for the high and low roles are presented in Table 1. In order to verify the assumption that respondents in the high role recordings would be perceived as

[3] Since intelligence could be related to the ability to use information for prediction purposes, partial correlations among intelligence, self F score, and predicted F score were computed. For this sample, intelligence is not a crucial factor for the prediction of F scores of other persons.

TABLE 1
COMPARISON OF MEAN ESTIMATED F SCORES FOR HIGH AND LOW ROLES

Role on Record	Subject Group	Information Session	N	Mean	SD	t	df[a]	p
High	Total	I	48	138.1	12.60	12.38	77	<.001
Low	Total	I	48	110.8	21.43			
High	Total	II	48	168.4	12.37	18.17	74	<.001
Low	Total	II	48	103.0	12.81			
High	High	I	24	157.7	11.94	9.39		<.001
Low	High	I	24	123.6	12.67			
High	High	II	24	167.8	11.64	13.77		<.001
Low	High	II	24	117.8	12.94			
High	Low	I	24	152.4	12.69	10.73	39	<.001
Low	Low	I	24	98.0	20.79			
High	Low	II	24	168.9	13.04	16.97	43	<.001
Low	Low	II	24	88.3	18.67			

[a] Degrees of freedom equal $N_1 + N_2 - 2$. Where variances are unequal,

$$df = \frac{n\left(\frac{(s_1)^2}{N_1} + \frac{(s_2)^2}{N_2}\right)}{\frac{(s_1^2)^2}{N_1}\frac{1}{N_1+1} + \frac{(s_2^2)^2}{N_2}\frac{1}{N_2+1}} - 2$$

(from Walker & Lev, 1953, p. 158)

scoring higher in F than those in the low role recordings, the roles were compared for the total subject group for each information session. The high role was, on the whole, differentiated from the low role for both information sessions. A comparison of *Ss* high in F judging the high role and *Ss* high in F judging the low role indicates that *Ss* high in F were able to differentiate between high and low roles for both information sessions. A similar comparison for *Ss* low in F resulted in the same finding.

The mean difference between the high role judgment and the low role judgment for *Ss* high in F was compared with that for *Ss* low in F. Both after Information Session I and after Session II, *Ss* low in F perceived greater differentiation between the two roles than did *Ss* high in F (I: $t = 3.25$, $p < .01$; II: $t = 5.12$, $p < .001$). The first hypothesis is thus supported.

High Subjects versus Low Subjects

Ss high in F did not differ significantly from *Ss* low in F in their judgments of the high role recordings following Information Session I ($t = 1.46$) or Information Session II ($t = .30$). On low role recordings, however, *Ss* low in F perceived the role as significantly lower than did *Ss* high in F following Information Session I ($t = 5.05$, $p < .001$) and Information Session II ($t = 6.24$, $p < .001$). The difference between *Ss* high and low in F on predictions for the low role was significantly greater than that for the high role after Information Session I ($t = 3.25$, $p < .01$) and after Information Session II ($t = 4.75$, $p < .001$). Following both information sessions *Ss* low in F thus perceived the low role more accurately than did *Ss* high in F. The two groups of *Ss* did not differ in their perceptions of the high role.

Information Session I versus Information Session II

A comparison of Information Session I and Information Session II for *Ss* high in F who judged the high role reveals that predictions were significantly more accurate following Information Session II ($t = 2.90$, $p < .01$). For the low role, *Ss* high in F were not significantly more accurate after Information Session II ($t = 1.53$). Similar results were obtained for *Ss* low in F; the high role was perceived more accurately after Information Session II than after Information Session I ($t = 4.35$, $p < .001$), but there was no significant difference with respect to the low role ($t = 1.67$).

Perhaps these findings might be influenced by ceiling effects. The highest judgment one could make for the high role was a score of 196. The lowest judgment one could offer for the low role was a score of 28. The more accurate a prediction was after Information Session I, the less it could be improved after Information Session II because of these set limits. To remove this inequity, improvement was rated relative to the possible improvement that could have occurred between the end of Information Session I and the end of Information Session II. Improvement was greater for *Ss* low in F than for *Ss* high in F for the high role ($t = 2.14$, $p < .05$). Improvement on the low role did not differ for *Ss* high and low in F ($t = 1.07$).

TABLE 2
MEAN SUMS OF DIFFERENCES BETWEEN SELF-RANKINGS OF VALUES AND RANKINGS OF ASCRIBED VALUES

Role on Record	Subject Group ($N = 24$)	Information Session	Mean	SD	t	df	p
High	High	I	8.0	4.04	4.85	36	<.001
High	Low	I	12.8	2.44			
Low	High	I	10.1	4.98	.51		—
Low	Low	I	9.4	3.76			
High	High	I	8.0	4.04	1.33		—
Low	High	I	10.1	4.98			
High	Low	I	12.8	2.44	3.58	40	<.001
Low	Low	I	9.4	3.76			
High	High	II	9.6	4.16	4.00	38	<.001
High	Low	II	13.6	2.38			
Low	High	II	8.0	3.74	.55		—
Low	Low	II	8.7	4.42			
High	High	II	9.6	4.16	1.35		—
Low	High	II	8.0	3.74			
High	Low	II	13.6	2.38	6.68	37	<.001
Low	Low	II	8.7	4.42			

Discrepancy between Own and Ascribed Values[4]

The means of the sum of differences between self-rankings of values and rankings of values ascribed to respondents in the recordings are presented in Table 2. *S*s low in F saw a greater discrepancy between their own values and those of respondents in the high role than did *S*s high in F following each information session. The two groups did not differ significantly in their discrepancy scores for the low role after either information session. *S*s high in F did not discriminate between ascribed values for the high and low roles following either information session, whereas *S*s low in F saw a greater discrepancy between self-values and values of the high role than between self-values and values ascribed to the low role after each information session. The second hypothesis is supported.[5]

DISCUSSION

The data on judgments of F scale scores show that *S*s high in F are able to make social discriminations when information is sufficiently relevant. It is not believed that the present finding represents evidence contradictory to the findings of previous studies (Crockett & Meidinger, 1956; Scodel & Mussen, 1953) that *S*s high in F distort information to satisfy a need to see others like themselves. But there seem to be definite limits within which distortion can occur. When information is as direct and obvious as in this study, *S*s high in F take it into account. However, their ability to discriminate is not as great as that of *S*s low in F. The advantage of *S*s low in F is due to their ability to perceive the low role more accurately. Even when information is highly relevant, the need of *S*s high in F to see others like themselves persists to some degree.

It has been noted that the present findings indicate that *S*s low in F differentiated between the two roles even more sharply than *S*s high in F. In previous studies (Crockett & Meidinger, 1956; Scodel & Freedman, 1956) *S*s low in F tended to place others in a middle category. The results obtained here indicate that *S*s low in F, given relevant information, are able to make accurate social judgments. Yet Crutchfield's (1954) suggestion that *S*s low in F may have an "other" stereotype (rather than the "same" stereotype of *S*s high in F) still seems valuable. An inspection of the raw data reveals that it is a rare *S* low in F who judges the scores for the low role as lower than his own (two of 24 after Information Session I; five of 24 after Information Session II). Similarly, Rabinowitz (1956) found a marked tendency among nonauthoritarians to estimate F scale scores for the typical college student as higher than they themselves possessed.

The inability of *S*s low in F to benefit from the additional information of Session II could be due to the role script itself. It could also be

[4] In a comparison of differences in self ranking of values for the two groups, *S*s high in F ranked authoritarianism significantly higher than did *S*s low in F (chi square value 7.19, $p < .01$). *S*s low in F ranked Integrity as significantly higher than did *S*s high in F (chi square value 7.43, $p < .01$).

[5] To extend the dimension of relevance, judgments were also sought on variables (height and weight) for which the information provided was in no way relevant. Analyses of discrepancies between self height and weight and judged height and weight resulted in generally negative findings. The one significant comparison indicated that *S*s low in F judged the low role as more similar to themselves in height and weight than the high role. For a fuller discussion, the reader is referred to the author's dissertation (Lipetz, 1958).

due to their own possible holding of irrational stereotypes. If, for the most part, they see others as having higher F scale attitudes than their own, they have already set an irrational lower limit for the low role; hence, they are unable to use the additional information.

The analyses involving value discrepancies between self and judgments offer some interesting findings. There can be little doubt that the information on the recordings was relevant for the ranking of values. The relevance, however, was not as direct as for the F scale judgments, since the scripts were written specifically with F content in mind. With reality factors less compelling, *Ss* low in F were still sensitive to differences between the roles. The data obtained from *Ss* high in F, however, are consistent with the Scodel-Mussen (1953) assertion that the person high in F maintains his security by assuming that the values of others are similar to his own.

In conclusion, it seems that both *Ss* high and low in F are governed to some extent by stereotyped thinking, albeit of different types.

SUMMARY

This study was concerned with the effects of differentially relevant information on social judgments of authoritarians and nonauthoritarians. After listening to standard recorded interviews, the content of which was derived from the theoretical clusters of the F scale, *Ss* made judgments about the respondent's F scale attitudes and his values.

It was hypothesized that *Ss* low in F are better able to use information for making accurate judgments. On the whole, results confirmed the hypothesis, although there was some evidence that *Ss* low in F, as well as those high in F, tended toward stereotyped thinking.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

CHRISTIE, R., & COOK, PEGGY. A guide to published literature relating to the authoritarian personality through 1956. *J. Psychol.*, 1958, **45**, 171–199.

CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 1956, **53**, 378–380.

CRUTCHFIELD, R. S. Social psychology and group processes. *Annu. Rev. Psychol.*, 1954, **5**, 171–202.

JUUL, K. D. Authoritarian personality in relation to teachers' attitudes toward child behavior. Unpublished doctoral dissertation, Wayne Univer., 1953.

KOGAN, N. Authoritarianism and repression. *J. abnorm. soc. Psychol.*, 1956, **53**, 34–37.

LIPETZ, M. E. The effects of information on the assessment of attitudes by authoritarians and nonauthoritarians. Unpublished doctoral dissertation, Ohio State Univer., 1958.

RABINOWITZ, W. A note on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **53**, 384–386.

SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92–95.

SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.

WALKER, HELEN M., & LEV, J. *Statistical inference.* New York: Holt, 1953.

WERTHEIMER, M. Values in person cognition. In USAF Behav. Sci. Conf. Rep., Univer. New Mexico, 1957, *Decisions, values, and groups.* New York: Pergamon Press, in press.

(Received January 7, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SOCIALLY MEDIATED REDUCTION IN EMOTIONAL RESPONSES OF YOUNG RHESUS MONKEYS[1]

WILLIAM A. MASON[2]
University of Wisconsin

CHILDREN frequently become upset and fearful when alone in strange surroundings (Arsenian, 1943; Jersild & Holmes, 1935), and nonsystematic observations of young monkeys and chimpanzees suggest that isolation in an unfamiliar situation produces similar effects in these nonhuman primates (Finch, 1942; Foley, 1934; Haslerud, 1938; Köhler, 1927; McCulloch, 1939).

Arsenian (1943) has demonstrated experimentally with preschool children that emotional distress induced by placing the child alone in a novel room is substantially reduced if a familiar adult is present. Similarly, the presence of familiar social stimuli has been observed to mitigate the emotional reactions of nonhuman primates to novel or frustrating situations (Haslerud, 1938; McCulloch, 1939; Watson, 1914).

Although it is reasonable to assume that the effectiveness of a specific social stimulus in reducing disruptive emotional responses is dependent upon the nature and extent of previous contacts with it, this assumption has apparently not been subject to systematic investigation. For most immature primates, the mother is no doubt highly effective in reducing disruptive emotional responses, but there is evidence that other adults (Arsenian, 1943) or children (Wright, 1942) have a similar effect. Interpretation of this finding is difficult, however, in view of demonstrated interaction between experience with the mother and subsequent responses to other social stimuli (Bishop, 1951).

The present subjects (*Ss*) provided a large measure of control over possible interactions of this sort. They were young monkeys, born in the laboratory and raised under conditions which, with the specific exceptions described below, limited social experience with other monkeys to visual and auditory contacts with age-peers. The laboratory environment provided everything essential to maintain the *Ss* in good health, but in a routine and impersonal manner. Thus their social history severely restricted opportunity for the development of complex social interactions and expectancies, particularly in regard to other monkeys.

The present researches investigated the influence of various social stimuli on the emotional responses of these *Ss* by testing them in isolation and in the presence of familiar and unfamiliar social stimuli.

EXPERIMENT 1

Method

Subjects. The *Ss* were 12 young rhesus monkeys 250 to 263 days of age. They were born in the laboratory, separated from their mothers within 18 hrs. after birth, and subsequently housed in individual living cages which allowed them to see and hear other young monkeys, but restricted other forms of social contact. Opportunity for unrestricted social interaction was limited to 60 daily 3-min. free-situation tests with the same partner completed approximately one week before the present experiment began. These tests were conducted in the cubicle subsequently described. Each *S* had had five additional 3-min. periods alone in the cubicle in the week preceding the present experiment.

Apparatus. The test chamber was a 6 ft. by 6 ft. by 6 ft. plywood cubicle illuminated from above by a 150 w. bulb. Two opposing walls of the chamber contained one-way observation panels 15½ in. by 11½ in. and in each of the other walls was a large hinged door fitted with a smaller sliding door through which the *S* entered and left the chamber. Only one sliding door was used in the present experiment. Near the wall opposite this door was a Plexiglas detention cage measuring 18 in. high, 12 in. wide, and 12 in. deep. The floor of the chamber was marked off into 2 ft. squares.

In addition to the detention cage, the only other objects within the cubicle were a red wooden cylinder 3 in. in diameter and 3½ in. high, a polished brass knob within a small box, enclosed on all sides but one, an unpainted wooden dumbbell 9 in. long, and a "tree" consisting of a wooden cylinder 3¼ in. in diameter and 15½ in. high into which several "limbs" of brass rodding were fixed. These objects were provided as incentives for visual and manual exploration. Two objects

[1] Support for this research was provided through funds received from the Graduate School of the University of Wisconsin, Grant G6194 from the National Science Foundation, and Grant M-722 from the National Institutes of Health.

[2] Now at Yerkes Laboratories of Primate Biology, Orange Park, Florida.

were placed near each of two walls on opposite sides of the room.

Procedure. At the beginning of each session the *S* was brought to the test area in a carrying cage and placed outside the cubicle in front of the sliding door. A trial was started by raising this door and introducing the *S* into the room. The 3-min. trial period was divided into 12 intervals each 15 sec. in length and the intervals were indicated by a Veeder-Root counter activated by a recycling timer. The *S* was observed through the one-way vision screens and its behavior recorded on a time-ruled check sheet. The recorded behaviors included the following emotional responses described by Foley (1934, 1935): (*a*) *Crouching-clasping.* Assumption of a crouching posture with limbs close to the body and head below the level of the shoulders and/or clasped its fur, limbs, or other body parts with hands or feet. (*b*) *Stereotyped movements.* Fixed, repetitive movements not resulting in locomotion. Rocking or swaying of the upper body were most frequently observed, but rubbing or stroking of body parts or contiguous regions of the chamber were also included in this category. (*c*) *Sucking.* Insertion into the mouth of fingers, toes, or other body parts. (*d*) *Vocalization.* Two forms of vocalization were recorded, a high-pitched, moderately intense "coo," and an intense scream or screech. A record was also obtained of *approaches to the detention cage, approaches to the objects* (the criterion of approach in each instance being 6 in.), and *contacts* with the objects.

Occurrence of any of the foregoing behaviors was recorded at the appropriate time interval on the check sheet, and scores for behavior categories were based on the number of 15-sec. intervals in which these behaviors occurred. No interval could contribute more than one to the total score for any behavior category regardless of the number of occurrences of the behavior within that interval. Locomotor activity was measured by advancing a counter whenever the *S* entered a square up to the midline of its body.

To provide an estimate of the reliability of the method, two *O*s were present and independently recorded responses for four *S*s on a total of 26 trials. Reliability measures were not determined for locomotion scores, and vocalization occurred too infrequently in these *S*s to warrant calculation of reliability coefficients. For all other measures, product-moment correlation coefficients ranged from .94 to .99.

The effects of three stimulus conditions were investigated: Each *S* was observed when the detention cage was empty (empty = Em), when the cage contained the partner that had been paired with the *S* in the 90 free-situation tests preceding the present experiment (partner = Pa), and when the cage contained an unfamiliar monkey of approximately the same age as the *S* (stranger = St). Stimulus conditions were presented in one of three fixed orders and all conditions appeared once before any condition was repeated. Each condition was presented five times.

Results

Total scores for each behavior category and their percentage equivalents are given for test conditions in Table 1. All emotion measures increased under the Em condition, but the effect of test conditions was statistically significant only for crouching-clasping and stereotyped movements. These measures were significant at the .01 level as determined by the Friedman nonparametric analysis of variance (Siegel, 1956).

TABLE 1
RESPONSES TO PARTNER, STRANGER, AND EMPTY STIMULUS CONDITIONS

	Partner		Stranger		Empty	
	Freq.	%	Freq.	%	Freq.	%
Crouching-clasping	265	37	264	37	409	57
Stereotyped movements	140	19	127	18	226	31
Sucking	68	9	56	7	86	12
Vocalization	15	2	10	1	31	4
Composite score	488	17	451	16	752	26
Aproach cage	601	83	611	85	425	59
Approach objects	223	31	265	37	239	33
Object-contact	182	25	205	28	182	25
Mean locomotion/trial	24.6		27.7		22.9	

Emotion scores were additively combined to yield a composite emotion score. The percentage of composite emotion scores were 17, 16, and 26 for the Pa, St, and Em conditions respectively, and the difference among conditions was significant at the .01 level. Comparisons between individual test conditions with the Wilcoxon test (Siegel, 1956) indicated that the Em condition differed from the Pa and St conditions at the .01 level. The difference between the Pa and St conditions was not statistically significant.

Approaches to the cage dropped from approximately 84% under the Pa and St conditions to 59% under the Em condition, and the effect of stimulus conditions on this measure was significant at the .01 level as determined by the Friedman test. Locomotion scores were also lowest under the Em condition, but differences among conditions were not statistically significant. There was no indication that approaches to objects or contacts with objects were systematically related to stimulus conditions.

The present experiment demonstrated that presence of a social stimulus produced a reduction in emotional responsiveness, but provided no evidence that the specific characteristics of the social stimulus and *S*s' previous experience with it were effective variables. Therefore, another experiment was conducted that used a wider range of social stimuli.

EXPERIMENT 2

Method

Subjects. The six pairs of rhesus monkeys used in Experiment 1 served again as *S*s. Their ages ranged from 439 to 536 days. In connection with another experiment, six of the monkeys had been housed with their partners and had been living together for 18 weeks when the present experiment began. For the other *S*s, living conditions were as previously described. Following Experiment 1, opportunities for social contact with the partner were provided to all *S*s in 20 social preference tests, each 2 min. long and conducted in the test

TABLE 2
RESPONSES TO PARTNER, STRANGER, ADULT, RABBIT, AND EMPTY STIMULUS CONDITIONS

	Partner		Stranger		Adult		Rabbit		Empty	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Crouching-clasping	137	19	171	24	229	32	269	37	276	38
Stereotyped movements	70	10	72	10	124	17	143	20	124	17
Sucking	65	9	54	7	91	13	106	15	103	14
Vocalization	2	0	12	2	24	3	36	5	14	2
Composite score	274	9	309	11	468	16	554	19	517	18
Approach cage	671	93	654	91	497	69	557	77	581	81
Approach objects	113	16	101	14	173	24	108	15	97	13
Object-contact	100	14	92	13	141	20	101	14	84	12
Mean Locomotion/trial	36.0		32.6		26.3		24.2		29.1	

cubicle previously described, and in tests of food-sharing and communication. Only the tests of social preference permitted unrestricted social interaction.

Apparatus. The cubed room was again used and the experimental situation was unchanged, except that vertically striped panels were suspended from two walls of the test chamber and the original objects were replaced by four new stimulus objects: a coffee can, a grey rubber hemicylinder, a small aluminum-painted block with dowels protruding from it, and an orange and red "tree." The dimensions of these objects were approximately the same as those described in Experiment 1.

Procedure. Testing and recording procedures and behavioral measures were the same as those used in Experiment 1, except for the addition of two stimulus conditions bringing the total number to five: Each *S* was observed when the detention cage was empty (Em), when the cage contained the partner (Pa), an unfamiliar monkey of the same age (St), an unfamiliar adult (Ad), and a large albino rabbit (R). The *S*s had never before seen a rabbit and had seldom if ever seen adult monkeys. All stimulus conditions appeared once before any condition was repeated and each condition was presented five times.

Results

The data were first examined for differential effect of living arrangements on responses to the Pa and St conditions. The *S*s living continuously with the partner gave no evidence that the partner was more effective in reducing emotional responses than a stranger of the same age. For all *S*s differences between the Pa and St conditions were slight, and there were no significant differences between paired and solitary *S*s in response to these stimulus conditions. In view of these findings the data from all *S*s were pooled for further analysis.

Total scores for each behavior category and their percentage equivalents are given for test conditions in Table 2. Although emotion scores were lower under the Pa and St condition, the effect of stimulus conditions was significant only for crouching-clasping. Emotion scores were additively combined to yield a composite emotion score. The percentage of composite emotion scores for the Pa, St, Ad, R, and Em conditions were 9, 11, 16, 19, and 18, respectively, and the effect of stimulus conditions on this measure was significant at the .01 level as determined by the Friedman test. Comparisons between individual stimulus conditions with the Wilcoxon test indicated that the Pa condition differed from the Ad, R, and Em conditions at the .02 level. The difference between the St and R conditions was also significant at the .02 level. All other differences were nonsignificant.

The percentage of approaches to the cage were 93, 91, 81, 77, and 69 for the Pa, St, Em, R, and Ad conditions, respectively, and the difference among conditions was significant at the .01 level as determined by the Friedman test. Comparisons between stimulus conditions indicated that the Ad condition differed from the Pa, St, and Em conditions at the .05 level as determined by the Wilcoxon test. All other differences were nonsignificant. Mean locomotion scores were highest in the Pa and St conditions, but differences among conditions were not significant for this measure or for approaches to objects and contacts with objects.

Approximately 6 months separated Experiments 1 and 2. The absolute magnitude of composite emotion scores in Experiment 2 was somewhat lower than for comparable stimulus conditions in Experiment 1, suggesting that the intervening experience in the test chamber and/or maturational changes had produced some reduction in the general level of emotional responsiveness. Despite this change, however, a rank order correlation performed on total individual composite emotion scores yielded a coefficient of .74 ($p = .02$), indicating that relative to other members of the group, individual response characteristics were stable and persistent.

DISCUSSION

Previous observations that social stimuli may function as a source of security and a means of mitigating emotional distress in young primates are fully supported by the present results. The assumption that the familiarity of the social stimulus is an important determinant of the capacity of the

stimulus to reduce emotional distress is confirmed by the finding that the level of emotional responses was lower in the presence of monkeys the same age as the *S* than in the presence of an unfamiliar adult monkey or rabbit. This conclusion must be qualified, however, in view of the failure to demonstrate unequivocally that the monkey with which the *S* had had the greatest amount of previous contact was more effective than an unfamiliar monkey of the same age. We possess unpublished data showing that the capacity for social discrimination is highly developed in monkeys of this age, and it is unlikely that failure to discriminate can account for the apparent equivalence of the familiar partner and the strange age-peer. A more acceptable hypothesis is that as the result of visual presence of young monkeys in the *S*'s home environment, all young monkeys had acquired equivalent status as familiar social stimuli in this situation. The effectiveness of the strange age-peer may also have been enhanced, however, as a result of the *S*'s experience with the familiar partner.

Although the present experiments did not investigate the effectiveness of familiar nonsocial stimuli, the research of Harlow and Zimmermann (1958) indicates that equivalent exposure in the living cage to two stimuli, one a preferred social substitute and the other nonpreferred, does not produce equivalent effects. Their research clearly demonstrates that the capacity of a stimulus to mitigate emotional distress is influenced by the specific characteristics of the stimulus in addition to the amount of previous exposure to it.

These considerations, and the obvious similarity of young and adult monkeys, suggest that the present *S*s should have reacted differently to the adult monkey than to the rabbit or the empty cage. Further information as to the possible nature of these differences is provided by Hebb's (1946) theoretical analysis of fear behavior. In his discussion of fear of strangers in young chimpanzees, Hebb proposes that a stimulus which displays sufficient similarity to familiar stimuli to arouse habitual processes of perception, but is also different in certain respects, will produce emotional disturbance. Thus, the adult monkey which was highly similar to the young monkeys with which these *S*s were familiar, but also differed conspicuously from young monkeys in many respects, should have produced emotional disturbance. Although composite emotion scores were actually somewhat lower in the presence of the adult monkey than in the presence of the rabbit or empty cage, the hypothesis that the adult monkey was a source of emotional disturbance is supported by the finding that approaches to the cage were lowest when the adult monkey was present and by the observation that stimulus-directed social responses including threat reactions, lipsmacking, and grimaces (usually interpreted as indicative of fear or submission) were most frequent for this stimulus condition. These responses were recorded systematically but were not included in the presentation of results because they are characteristically elicited by and directed toward social stimuli (Bernstein, 1959; Carpenter, 1942; Chance, 1956), in contrast to the composite emotion measures which occur in the absence of a discrete eliciting stimulus (Bernstein, 1959). Threat reactions, lipsmacking, and grimaces occurred approximately twice as frequently in the presence of the adult monkey as compared with any of the other stimulus conditions, and significant differences were obtained between this stimulus condition and the rabbit and empty cage. These results suggest that the adult monkey had an effect on the emotional responses of these young monkeys similar to that evoked in children and young chimpanzees by strange persons (Hebb & Riesen, 1943; Jersild & Holmes, 1935).

Contemporary theoretical discussions of social development of the human child emphasize the fundamental importance of nurturant activities, particularly feeding experiences, in the development of social attitudes and patterns of social interaction (Dollard & Miller, 1950; Mussen & Conger, 1956; Sears, Whiting, Nowlis, & Sears, 1953). It is highly improbable that the effectiveness of the partner and the strange age-peer in reducing behavior indicative of emotional distress was in any manner derived from or dependent upon nurturant activities. The *S*'s only experience with other monkeys prior to Experiment 1 was confined to 90 brief trials with a single partner and the restricted contacts provided in the

living area. From birth, all major caretaking acts were performed routinely and impersonally by laboratory personnel. Although performance of nurturant functions might substantially enhance the capacity of a social stimulus to reduce emotional disturbance, it is clear from the present results that it is not a necessary antecedent of this effect.

SUMMARY

Two studies investigated the effect of various social stimuli upon emotional responsiveness of 12 young rhesus monkeys raised in the laboratory in the visual presence of other young monkeys. Experiment 1 compared the effect of three stimulus conditions: A familiar partner, an unfamiliar age-peer, and the empty detention cage, and Experiment 2 extended the range of test stimuli to include an adult monkey and an albino rabbit.

In both experiments responses indicative of emotional disturbance were less frequent in the presence of the partner and the unfamiliar age-peer. These results suggest that the nature of *Ss*' previous social experience was an important determinant of stimulus effectiveness in reducing emotional distress. The capacity of a social stimulus to mitigate emotional disturbance did not appear to be dependent upon nor derived from feeding or other nurturant experiences. Evidence was presented suggesting that the adult monkey was a source of emotional disturbance, producing an effect similar to that observed in young children and chimpanzees in the presence of strangers.

REFERENCES

ARSENIAN, JEAN M. Young children in an insecure situation. *J. abnorm. soc. Psychol.*, 1943, **38**, 225–249.

BERNSTEIN, S. The development of emotional responses to complex stimuli in young rhesus monkeys. Unpublished master's thesis, Univer. of Wisconsin, 1959.

BISHOP, B. M. Mother-child interaction and the social behavior of children. *Psychol. Monogr.*, 1951, **65**, (11, Whole No. 328).

CARPENTER, C. R. Sexual behavior of free ranging rhesus monkeys (*Macaca mulatta*): I. Specimens, procedures, and behavioral characteristics of estrus. *J. comp. Psychol.*, 1942, **33**, 113–142.

CHANCE, M. R. A. Social structure of a colony of *Macaca mulatta*. *Brit. J. Anim. Behav.*, 1956, **4**, 1–13.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

FINCH, G. Chimpanzee frustration responses. *Psychosom. Med.*, 1942, **4**, 233–251.

FOLEY, J. P., JR. First year development of a rhesus monkey (*Macaca mulatta*) reared in isolation. *J. genet. Psychol.*, 1934, **45**, 39–105.

FOLEY, J. P., JR. Second year development of a rhesus monkey (*Macaca mulatta*) reared in isolation during the first eighteen months. *J. genet. Psychol.*, 1935, **47**, 73–97.

HARLOW, H. F., & ZIMMERMANN, R. R. The development of affectional responses in infant monkeys. *Proc. Amer. Phil. Soc.*, 1958, **102**, 501–509.

HASLERUD, G. M. The effect of movement of stimulus objects upon avoidance reactions in chimpanzees. *J. comp. Psychol.*, 1938, **25**, 507–528.

HEBB, D. O. On the nature of fear. *Psychol. Rev.*, 1946, **53**, 259–276.

HEBB, D. O., & RIESEN, A. H. The genesis of irrational fears. *Bull. Canad. psychol. Ass.*, 1943, **3**, 49–50.

JERSILD, A. T., & HOLMES, FRANCES B. *Children's fears*. New York: Teachers Coll. Columbia Univer., 1935.

KÖHLER, W. *The mentality of apes*. New York: Humanities Press, 1927.

MCCULLOCH, T. L. The role of clasping activity in adaptive behavior of the infant chimpanzee: III. The mechanism of reinforcement. *J. Psychol.*, 1939, **7**, 305–316.

MUSSEN, P. H., & CONGER, J. J. *Child development and personality*. New York: Harper, 1956.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, **47**, 135–236.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

WRIGHT, M. E. Constructiveness of play as affected by group organization and frustration. *Character & Pers.*, 1942, **11**, 40–49.

WATSON, J. B. *Behavior: An introduction to comparative psychology*. New York: Holt, 1914.

(Received January 12, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

DIRECT AND VICARIOUS REDUCTION OF HOSTILITY[1]

MILTON E. ROSENBAUM
State University of Iowa

AND RICHARD DECHARMS
Washington University

OBSERVATION of the responses of another to various environmental stimuli may at times affect the response system of the observer as if he had carried out the behavior himself. Phenomena of this type are described as falling under the rubric of vicarious experience (deCharms & Rosenbaum, in press). A minimal amount of research attention has been given to vicarious phenomena, although many students of the social scene claim increasing participation in this mode of experience in contemporary society.

The present report examines the effectiveness of vicarious experience in reducing hostility toward an instigator and at the same time tests hypotheses related to the role of communication in the reduction of aggressive responses.

The theory of frustration and aggression as elaborated by Dollard, Doob, Miller, Mowrer, and Sears (1939) posits that the blocking of goal directed activity leads to the occurrence of aggression of some kind and in some degree. They further postulate in a dependent definition of aggression that aggression "reduces the secondary, frustration-produced instigation, and leaves the strength of the original (nonaggressive) instigation unaffected" (Dollard et al., 1939, p. 11). The statement concerning the reduction of the secondary instigation represents the catharsis hypothesis. More specifically, the catharsis hypothesis reads: "The occurrence of any act of aggression is assumed to reduce the instigation to aggression" (p. 50).

Dollard et al. present minimal evidence in support of the hypothesis. Several later investigators have treated the catharsis hypothesis in contexts involving displacement of aggression. Reduction of experimentally aroused hostility has been studied most frequently as a function of aggression expressed toward objects other than the original instigating agent, i.e., a TAT card (Feshbach, 1955) or associates (French, 1944; Thibaut, 1950; Pepitone & Reichling, 1955). In general, support for the hypothesis was obtained in these studies, although as Berkowitz (1958) indicates, alternative explanations are available in each case.

In two experiments involving expression of aggression toward the instigator himself, Thibaut and Coules (1952) obtained equivocal results related to the catharsis hypothesis. Their basic design involved a situation where, in the course of interaction, *S* was personally insulted in a note sent by a paid participant. Half of the *S*s were allowed to respond immediately, while for the other half the interaction was interrupted at that point. A measure of residual hostility gave results in the direction predicted by the catharsis hypothesis, i.e., those *S*s who were permitted to communicate their hostility directly to the aggressor showed less residual hostility.

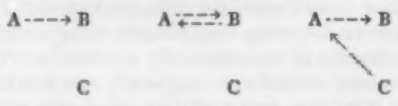
As Thibaut and Coules (1952) pointed out, this finding may have been due to increased frustration produced by the blocking of a hostile response in the interrupted group rather than to a genuine catharsis effect. Experimental evidence was obtained supporting this possibility. In addition, there was no indication within the communication group that the amount of expressed aggression was related to the amount of reduction in hostility, as would have been predicted by the catharsis hypothesis. An implicit assumption of equal instigation is necessary in the examination of

[1] This paper is based on work carried on while the authors were in attendance at the Behavioral Sciences Research Conference at the University of New Mexico during the summers of 1957 and 1958. This conference was supported by the Behavioral Sciences Division, Air Force Office of Scientific Research, under Contract AF 49(6380-33). A portion of the final data analysis was carried out under Contract Nonr-816(11) between the Office of Naval Research and the Small Groups Laboratory, Social Science Institute at Washington University. The authors are indebted to I. E. Farber for a critical reading of the manuscript.

this relationship. Evaluation of their own data by Thibaut and Coules indicates that this assumption was not fulfilled.

By introducing a third person into an experimental situation similar to that of Thibaut and Coules, it is possible to gain further insight into the dynamics underlying their results and also to explore conditions under which vicarious experience is successful in modifying responses.

Diagrammatically, three experimental conditions can be represented as follows:



1. No Communication (No-Com) | 2. Communication (Com) | 3. Vicarious Communication (Vi-Com)

In these diagrams the initial aggressor is always represented by A, the *S* who is aggressed against is represented by B, and the potential "vicar" is represented by C. The arrows represent aggressive attacks, or, in the communication situation, the possibility for B to retaliate against A by means of an aggressive communication to him. Situations 1 and 2 are, in effect, those investigated by Thibaut and Coules with the addition of the bystander, C. Situation 3 represents a vicarious setting in which C acts as a "vicar" for B.

An important variable to be accounted for in this model is the strength of the original instigation to aggression. The actual insulting attack, or in terms of the Dollard et al. (1939) formulation, the degree of interference, can be constant across experimental conditions, but individual reactions to the attack may vary. Since instigation to aggression is conceived of as being an interference with goal directed activity, strength of instigation to aggression will be a function of the strength of the instigation to the activity that is interfered with. In the experiment to be described, the activity is the presentation of information about oneself. This task may be of varying importance to the person in the situation, depending on his level of self-esteem. The person with low self-esteem who feels socially inadequate can be expected to be more concerned with presenting an acceptable impression and receiving favorable evaluation than the person with high self-esteem; and, therefore, the attack on the self-picture of the person low in self-esteem may engender more frustration and hence more aggression.

The experiment to be reported was designed to study the residual hostility shown by *S*s after experiencing one of the three conditions in the model. Specifically, the hypotheses to be tested are:

Hypothesis I. A person will manifest less residual hostility after an aggressive attack on him if he is given an opportunity to communicate directly to the aggressor than if such an opportunity is denied.

Hypothesis II. A person will have less residual hostility after an aggressive attack on him if someone else within his awareness attacks the aggressor than if no opportunity for him or anyone else to attack is presented.

The effectiveness of the vicarious situation for reducing hostility will depend upon characteristics of the vicarious attack and upon characteristics of the vicar. Despite these factors one would predict that direct opportunity for attack should be more effective than a vicarious attack in reducing hostility. In effect, we are presenting as a tentative hypothesis the contention that the amount of residual hostility in the vicarious situation should be intermediate between the amounts in the other two situations.

Hypothesis III. A person with low self-esteem will manifest the effects predicted in Hypothesis II and III to a greater extent than will a person with high self-esteem.

METHOD

Subjects

Fifty-six male summer session students at the University of New Mexico participated in the experiment. The *S*s were paid for their two hours of participation. Four of these *S*s spontaneously indicated suspicion of deception and were therefore not considered in the final analysis. Two other *S*s failed to complete all of the experimental forms. A total of 50 *S*s comprised the experimental sample.

Equipment

The experiment was conducted in four small semi-soundproof rooms, one of which served as a master control room and the others served as experimental rooms in which the *S*s were isolated individually. It was possible, therefore, to have a maximum of three *S*s at one time. This number was optimal, since every

effort was to have *S* believe that he was one of a group of three. Care was taken, however, to insure that *S*s did not see each other before the experiment, and it was possible to conduct the experiment even when only one *S* reported.

Each experimental room contained a desk and chair, and *S* was provided with a pencil, a set of headphones, and a microphone. The master control room contained four tape recorders. Separate recorders were available to record and/or monitor everything that happened in each of the three experimental rooms. The fourth recorder played a master tape that was heard over the headphones in all three of the experimental rooms.

Procedure

As soon as *S* was ushered into his experimental room, he was handed a form entitled Value Profile. This contained 20 statements which the *S* rated as to his agreement or disagreement on a seven-point scale. Ten of the items comprised the Self-Esteem scale which measures feelings of social inadequacy. These items were adapted from items used by Janis (1954). They had been previously shown to be highly intercorrelated and related to a conformity measure (Rosenbaum & deCharms, 1957). Ten items from the F scale were used as filler.

When *S* had finished the questionnaire, *E* gave instructions to each *S* individually concerning the remainder of the experiment, which was described as a study of the ability to evaluate others after a short nonvisual interaction. Since *E* visited each *S* individually, it was necessary for each *S* to sit quietly while the others were being instructed. The amount of time was roughly equated even in sessions when only one *S* was present.

The instructions prepared *S* to describe himself to two other participants following the topics and the order presented to him on the instruction sheet. This sheet made it clear to him when he was to speak and when the others were to speak. The sequence of speaking instructions included instances in which a single participant spoke to only one other person as well as occasions in which two others were addressed. The topics which each participant was to discuss were:

Round I. "Who are you?" This was to include relevant background information but no names.

Round II. Describe your social self.

Round III. Describe your nonsocial self.

Round IV. Characteristics that you dislike in yourself.

Round V. Characteristics that you like in yourself.

Round VI. Characteristics that you would like to possess ideally.

Brief explanations were given of each of these topics to give *S* a start. The instructions did not lead *S* to expect any spontaneous interaction with the other *S*s, since each was to be talking about himself and was not to ask or answer questions. Each *S* was told that he would be called B for the duration of the experiment. The people to whom he would be talking would be A and C. The order in which each spoke varied from round to round but the over-all order was constant for all conditions up to and including Round V. The order of speaking in Round VI varied according to the experimental condition.

TABLE 1
SEQUENCE OF INTERACTIONS IN ROUND VI

Condition	Minute 1	Minute 2 (The Attack)	Minute 3
No Communication	B to A & C	A to B & C	C to A
Vicarious Communication	B to A & C	A to B & C	C to A & B
Communication	C to A	A to B & C	B to A & C

A master tape which each *S* heard contained instructions from the *E* explaining the topic to be discussed and telling who was to speak to whom. The tape also contained standardized prerecorded speeches for both A and C for each of the rounds.[2] When B (the *S*) was speaking, the tape was silent and the other tape recorders recorded what each *S* actually said about himself on each round. Three versions of the master tape were used, one for each communication condition, and these were identical up to and including Round V. On Round VI, A, who had up to this time presented himself as a self-assured, rather egotistical person, attacked B (the *S*), calling him, among other things, an "egotistical, deceitful liar" and alleging that he was "full of bullshit." These phrases were taken from Thibaut and Coules (1952). This attack lasted for one minute as did all the other communication periods for each person within a single round. The attack always came in the middle of Round VI, i.e., A always spoke second on this round and spoke to both B and C (see Table 1). He directed his remarks completely to B, however. The experimental manipulation in the different versions of the tape on Round VI are presented in Table 1. In the No-Com condition, B spoke first to A, then A spoke to B and C (the attack on B), then B was not allowed to respond but had to sit in silence (as a time control) while C was supposed to be talking to A. In the Com condition, C spoke first to A while B sat in silence. Then A spoke to B and C (the attack), and then B was allowed to respond to A. In the Vi-Com condition B spoke first to A, then A to B and C (the attack) and then C attacked A (while B listened) for throwing his weight around, calling him "fantastic" and a "bastard" who "ought to live in a jungle." Care was taken that C did not mention B at all in his attack on A.

Immediately after Round VI, *E* visited each *S* individually, giving him a form on which he was to write a first draft of a personality sketch of both A and C, in that order, in which *S* was asked (*a*) to "Describe his personal characteristics as you see them," (*b*) "What characteristics of his do you like and/or dislike?" (*c*) "Under what circumstances, if any, and how frequently would you like to associate with this person?" (*d*) "Please give your over-all evaluation of him."

In order to increase the likelihood that *S* would perceive the personality sketch as a means for aggressing against the attacker, he was told that the person for

[2] We wish to thank Robert Gestaldo and Robert Morris for playing the roles on the master tape.

whom the form was filled out would be the only person to see it and that he would be allowed to make comments on it and return it to the *S*, at which time *S* would fill out the final draft to be turned in to *E*. In order to facilitate this procedure, *S*s were provided with envelopes addressed to A and C, and they enclosed their "first drafts" in the envelopes which were "delivered" by *E*.

When *S* had completed the personality sketch, the experimental session was essentially over. However, *E* visited each *S* individually, telling him that he had finished first and would have to wait to receive the ratings from the others. An interview was then carried out, starting with an inquiry about the electrical equipment and leading up to *S*s feelings about the situation. *E* attempted in every case (although he was not always successful) to get *S* to tell him about the attack, which *E* was not supposed to have heard. Once the attack was mentioned, *S* was encouraged to describe his feelings and was further pushed extensively to determine whether he had been at all suspicious of the situation. Once this information was obtained, the *S*s were brought together and, after the deception had been revealed, they were again asked about suspicion. The entire experiment was then explained in detail and rapport reestablished. Before leaving the experimental rooms *S*s were enjoined not to talk about the experiment. No *S* reported having heard anything about the experiment prior to appearing.

Analysis of the Data

The self-esteem score for each individual was the sum of his responses on a 7-point scale to the 10 self-esteem items on the pre-experimental value profile. Since half of these items were worded negatively, contamination by response set was avoided. High and low self-esteem scores were determined by a median split of all *S*s in the experiment. The distribution of scores was not skewed, and there were no significant differences between mean self-esteem scores in the three experimental treatments.

The primary dependent variable in the experiment was a residual hostility score derived from a content analysis of the "personality sketch." The responses on this form were first separated into thought units, usually phrases, and then each phrase was categorized as (*a*) positive evaluation of the person, (*b*) neutral, or (*c*) negative evaluation. Scoring was done by two scorers who knew nothing about the experiment.[3] Interscorer reliability was 92% agreement.

The score used as the measure of residual hostility was the percentage of thought units categorized as negative. Since there were very few neutral units and they were equally distributed, a score based on the positive units resulted in the same conclusions. There were no significant differences in mean total and number of units produced between experimental groups. "Residual hostility score" refers henceforth to the percentage of thought units categorized as negative in the personality sketch of the aggressor, A.

[3] Ruthe Harriger and Marvin F. Daley.

TABLE 2
MEAN RESIDUAL HOSTILITY SCORES (% NEGATIVE UNITS)

	No Communication	Vicarious Communication	Communication	Total
High self-esteem				
Mean	49.73	47.28	53.14	50.00
N	11	7	7	25
Low self-esteem				
Mean	68.28	49.60	42.63	52.60
N	7	10	8	25
Total				
Mean	56.94	48.65	47.53	
N	18	17	15	

RESULTS

Table 2 presents the residual hostility means by experimental conditions. The total means of the three communication conditions are in the order predicted by the hypotheses; namely, residual hostility is greatest in the No-Com condition, least in the Com condition, and intermediate in the Vi-Com condition. However, analysis of variance demonstrates that these differences are not significant. A test of the difference between the Com and No-Com conditions (Hypothesis I) reaches the .05 level of significance, confirming Thibaut and Coules' finding, although the test can be questioned since the over-all test does not reach significance.

Inspection of the body of Table 2 and of Fig. 1 makes it clear, however, that these tests, which ignore the level of self-esteem, obscure real differences. The difference apparent in Fig. 1 (solid lines) between the pattern of the high self-esteem means and that of the low self-esteem means in the evaluations of A is statistically significant.[4] It is obvious, therefore, that Hypothesis III is supported in the sense that the predicted effects occur most strongly in the low self-esteem groups. However, implicit in the hypothesis was the prediction that the effects would be similar but of a lesser magnitude for high self-esteem groups. This is clearly not supported. The high

[4] Both a test of interaction within an analysis of variance of the data in Table 2 and a test of the difference between the slopes presented in Fig. 1 reach statistical significance. The estimate of error derived from the analysis of variance was employed in further statistical tests when appropriate.

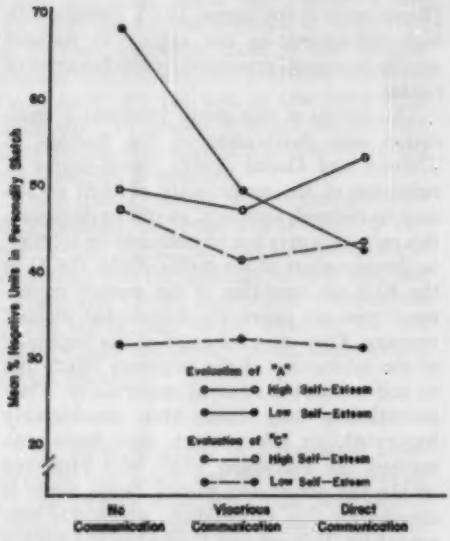


FIG. 1. Mean percentage of negative units in the personality sketches of "A" and "C."

self-esteem means are not in the order predicted and do not differ significantly. However, an analysis of variance of the three low self-esteem means demonstrates a significant difference ($p < .05$).

A more detailed analysis employing one-tailed t tests gives the following results:

Hypothesis I predicts a lower mean (less residual hostility) for the Com condition than for the No-Com condition. This hypothesis is supported for the *S*s with low self-esteem ($t = 2.61$, $df = 44$, $p < .01$).

Hypothesis II predicts a lower mean (less residual hostility) for the Vi-Com condition than for the No-Com condition. This hypothesis is supported for *S*s with low self-esteem ($t = 1.91$, $df = 44$, $p < .035$).

No significant difference can be demonstrated between the Vi-Com and the Com conditions. It is, therefore, impossible to show conclusively that the Com condition reduces hostility more efficiently than the Vi-Com condition, although the directional trend suggests that such is the case.

A subsidiary analysis of interest involves *S*'s evaluation of C. One might expect C to be more positively evaluated in the Vi-Com condition. In this situation, in contrast to the other two conditions, C behaved in a manner that was clearly negative toward A, although C was not at any time directly supporting of B (*S*). This should result in a more positive evaluation of C, especially for *S*s with low self-esteem, since it has been demonstrated that their responses to A were affected by the Vi-Com situation.

Content analysis of the personality sketches for C results in no support for this expectation (see Fig. 1, dotted lines). No statistically significant differences in evaluation of C can be demonstrated between the various communication conditions for either high or low self-esteem. There is, however, a significant difference ($p < .025$) between high and low self-esteem *S*s over all communication conditions, indicating that the low self-esteem *S*s (bottom line, Fig. 1) rate C more positively than do the high self-esteem *S*s in all conditions.

DISCUSSION

Hypotheses I and II predicted a greater amount of residual hostility in the No-Com condition than in either the Vi-Com or the Com conditions regardless of level of self-esteem. Hypothesis III predicted a difference between high and low self-esteem *S*s, but primarily in terms of relative differences in instigation to aggression. Since less instigation was expected for high self-esteem *S*s, it was expected that Hypotheses I and II would hold for these *S*s, but that the relationships found would not be as marked. The results support Hypotheses I and II unequivocally for *S*s with low self-esteem. Hypothesis III is supported in the sense that the results predicted by the first two hypotheses were indeed found most strongly with low self-esteem *S*s. No significant relationships were found, however, with high self-esteem *S*s. The failure of the predicted relationships to appear in the results for *S*s with high self-esteem must be accounted for.

Several possibilities for an explanation of the behavior of the high self-esteem *S*s are available. First, it is possible that these *S*s persist in hostile responses regardless of the means of expression available to them; i.e., after instigation to aggression no type of response opportunity presented in the experiment affects the level of residual hostility.

A counterproposal is that the high self-esteem *Ss* were not instigated to aggression, and the amount of residual hostility manifested is not related to the insult procedure. The evidence to follow would seem to favor the latter alternative. The assumption can be offered that residual hostility in the No-Com condition reflects a base line for amount of hostility, in that neither personal nor vicarious expression changes the level of residual hostility. Within the No-Com condition regression analysis indicates a significant difference ($p < .01$, 2 tails) between the high and low self-esteem *Ss*, the latter manifesting more residual hostility. This finding, which can be seen in Fig. 1, tends to indicate a lower level of instigation for the high self-esteem *Ss*, as was expected when the need for maintenance of self-esteem is the activity interfered with.

Since instigation was apparently low for *Ss* with high self-esteem, it is possible that they were not sufficiently instigated for aggressive responses to play an important role in the subsequently required personality sketches. This possibility would, of course, lead to lack of differences in residual hostility as a function of the communication conditions.

Some evidence can be brought to bear to support this possibility. Although no difference in evaluation of C in the different conditions was found, when a comparison is made of the difference in evaluation of A and C the results indicate that low self-esteem *Ss* present a significantly lower mean percentage of negative thought units in evaluating C than in evaluating A ($t = 3.50$, $p < .01$, 2 tails). The difference for the high self-esteem *Ss* although in the same direction is not significant ($t = .97$, $p > .30$, 2 tails).

It appears that the high self-esteem *Ss* in general perceived no difference between A and C, despite the fact that A under all conditions attacked them while C did not. This finding would seem to indicate a certain insensitivity to the comments of others on the part of high self-esteem *Ss*. The attacking statements made by A would appear to have failed to arouse anger in the high self-esteem *Ss*. This manifestation of insensitivity is consistent with other findings associated with the self-esteem variable. High self-esteem *Ss* have been found to be relatively nonresponsive to persuasion (Janis, 1954) and to efforts to elicit conformity (Rosenbaum & deCharms, 1957). Persons with high self-esteem do not appear to respond readily to stimuli presented by the behavior of others.

The results of this study represent a replication and clarification of the findings of Thibaut and Coules (1952). Some degree of reduction of the equivocality of their results may be claimed, although, as will be discussed, this reduction may not be complete. In contrast to the procedure of the earlier study, the *Ss* in the No-Com condition of the present experiment were not summarily interrupted without warning. They were informed at the beginning of the interaction of the sequence which was to end without a rebuttal opportunity. They nevertheless may have been considerably frustrated, an effect which may have augmented the aggression which was expressed within the personality sketch. Some doubt is created for this explanation, which was presented by Thibaut and Coules, in the finding of the present experiment that the Vi-Com *Ss*, who also were frustrated by lack of opportunity for self-expression, nevertheless present significantly less residual hostility than the No-Com *Ss*. It is, of course, possible that the augmented frustration still operates for the No-Com group and the Vi-Com situation serves successfully to relieve the frustration which accrued from lack of own expression. No way of deciding between these two alternatives is available in this experiment.

The catharsis hypothesis, which was the starting point for the Thibaut and Coules study and for the present experiment, apparently receives support from the results. It must be pointed out, however, that no direct evidence is available from either study to indicate the mechanism responsible for the lowered residual hostility in the Com condition as compared to the No-Com. In the present experiment there is scant evidence that aggressive verbalizations during the rebuttal are the cause of the reduced hostility in the Com condition. The rebuttal statements of the *Ss* were far milder than would be expected if a response in kind was forthcoming. That is to say, the general rebuttal response appeared far less aggressive than the standardized attack delivered by A. For example, although A employed profanity in his attack, none of the Com *Ss* responded in this manner. In fact,

clearly aggressive words were extremely rare and never approached the level of those used by A. Very frequently Com *Ss* said nothing relevant to the attack, but proceeded in Round VI to carry out the task of describing their "ideal self." In general, it seems unlikely that the verbalizations were of a character to lead to a reduction of hostility.

These results strongly suggest that a catharsis hypothesis, i.e., reduction of hostility due to expressed aggression, may be inadequate in accounting for the findings of this and the Thibaut and Coules experiment. Clarification of the mechanism which accounts for the hostility reduction implicit in this study awaits further research.

SUMMARY

Fifty *Ss* participated in a simulated interaction in which each was verbally attacked by a standard tape recording. Under three experimental conditions *Ss* (*a*) sat in silence after the attack, (*b*) heard a third person denounce the attacker, (*c*) were permitted to communicate back to the attacker themselves. *Ss* in each group were classified as to level of self-esteem. A measure of residual hostility toward the attacker constituted the primary dependent variable.

Results demonstrated (*a*) more residual hostility for low self-esteem *Ss* who were not allowed to reply to the aggressor than for similar *Ss* who could communicate or who heard someone else denounce the aggressor, (*b*) no significant differences between groups of high self-esteem *Ss*, (*c*) differences between the evaluation of the aggressor and of the third person for low self-esteem *Ss*, but not for high self-esteem *Ss*, and (*d*) very little actual aggression shown by any of the *Ss* who were allowed to communicate back to the aggressor after the attack.

These results give support to the hypotheses that both direct and vicarious communication to the aggressor reduces hostility for *Ss* with low self-esteem. It appears that high self-esteem *Ss* were not very strongly instigated to aggression, and that they showed some insensitivity to cues presented by other persons since they seemed not to discriminate between the aggressor and the other more benign person. An explanation of the results of this and of Thibaut and Coules (1952) study using a catharsis hypothesis was called into question, since very little actual aggression was observed in *Ss* response to the agressor, and yet this communication, which was in most instances unaggressive and unrelated to the attack, appears to have reduced hostility.

REFERENCES

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257–283.

DECHARMS, R., & ROSENBAUM, M. E. The problem of vicarious experience. In Dorothy Willner (Ed.), *Decisions, values, and groups*. New York: Pergamon, in press.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.

FRENCH, J. R. P. Organized and unorganized groups under fear and frustration. *Univ. Ia. Stud. Child Welf.*, 1944, **20**, 229–308.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

PEPITONE, A., & REICHLING, G. Group cohesiveness and the expression of hostility. *Hum. Relat.*, 1955, **8**, 327–339.

ROSENBAUM, M. E., & DECHARMS, R. Status revocability and social conformity. *USN ann. Tech. Rep.*, Nov. 1957, ONR855(04), Org. Res. Group, Univer. of North Carolina. Pp. 6–7.

THIBAUT, J. An experimental study of the cohesiveness of underprivileged groups. *Hum. Relat.*, 1950, **3**, 251–278.

THIBAUT, J. W., & COULES, J. The role of communication in the reduction of interpersonal hostility. *J. abnorm. soc. Psychol.*, 1952, **47**, 770–777.

(Received January 19, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

PREDICTION OF SMALL GROUP ROLE BEHAVIOR FROM PERSONALITY VARIABLES

WALTER R. BORG
Utah State University

THIS study explores the degree to which a person's role in a small group problem solving seminar can be predicted from a battery of group administered tests which attempt to measure skills and personality characteristics observed to be important in small group situations. The specific hypothesis to be tested is that a person's scores on tests selected to give measures of assertiveness, rigidity, self orientation, and sociability will predict his role in a problem solving seminar.

METHOD

Subjects

The *Ss* included 819 Air Force Majors and Lieutenant colonels in the Command and Staff School of the Air University, Class 1957. They were divided into 60 seminars of 12 to 14 officers each, and worked in the same seminar group for a period of two and one-half months. A large portion of the curriculum of the school is devoted to problem solving activities and discussions in the seminar groups. The problems dealt with by the seminars vary considerably in nature but, for the most part, involve practical Air Force problems of a sort that might be encountered by a Wing Staff.

Variables and Measures

The first step in the research was the selection of tests to go into the predictor battery, which was to contain tests to measure variables related to patterns of behavior that had been found in observations of small group activity. A careful survey was made in order to identify variables that had occurred repeatedly in studies based on observations of small group interaction (Bass, 1951; Bass, McGehee, Hawkins, Young, & Gebel, 1953; Carter 1951, 1954; Carter, Haythorn, & Howell, 1950; Carter, Haythorn, Shriver, & Lanzetta, 1951; Cattell & Stice, 1953; Haythorn, 1953; Sakoda, 1952; Schutz, 1955). The variables that appeared most frequently in these studies were:

Sociability. This variable, under different names, such as Personal Orientation, Group Sociability, and Social Adjustment, has been reported frequently in studies concerned with small group interaction.

Self orientation. Variations of this measure have appeared repeatedly under such names as Power Orientation (Schutz, 1955) and Individual Prominence, (Carter, 1954; Carter & Nixon, 1949).

Rigidity. Rigidity and related variables such as Authoritarianism, Conformity, and Dogmatism have been investigated in a number of relevant studies of small group behavior (Greer, 1955; Haythorn, Couch, Haefner, Langham, & Carter, 1956; Lippitt, 1940; McCurdy & Eber, 1954; Medalia, 1955).

Assertiveness. This variable has been employed with considerable success by Schutz (1955) and is somewhat similar to variables employed in other small group research such as Forceful Leadership, Surgency (Cattell, 1951), Ascendance (Bass et al., 1953), Physical Energy (Sakoda, 1952), and Individual Prominence (Carter, 1954).

The tests selected to supply predictive measures of these characteristics included the following:

The FIRO test. This test was developed by Schutz (1955) and has been used by him with considerable success for predicting leadership behavior and organizing compatible groups. The form of the test employed in this research consisted of seven items, each of which was regarded as a separate score on the suggestion of Schutz. These items attempt to measure degree of power orientation, as compared with personal orientation and degree of assertiveness.

The California F Scale, Form 40. This form of the F scale employs the items that were found to be most discriminating in the original research at the University of California (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950).

The Anxiety to Achieve Battery. This battery consisted of six group tests from the Cattell Objective Analytic Personality Battery (Cattell, 1955) and yielded a single factor score.

The Social Extroversion Score, from the Inventory of Factors STDCR (Guilford, 1945).

Eight factor scores from the Guilford Opinion Inventory (Guilford, 1951). This test was developed by Guilford from his factor analysis of the "interests" of Air Force officers. Factors used included: need for variety, orderliness, need for attention, physical drive, aggression, resistance to restriction, cultural conformity, self-reliance. The entire predictor battery yielded 18 scores.[1]

Derivation of Factor Scores

After the battery was administered, results were subjected to centroid factor analysis. A total of four orthogonal rotations were made, yielding four factors.

[1] Description data on the predictor battery and the six role descriptions used in the peer nominations have been deposited with the American Documentation Institute. Order Document No. 6094, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
ROTATED FACTOR LOADINGS EMPLOYED IN DEVELOPING WEIGHTED FACTOR SCORES

Variable	Loading
Factor 1: Assertiveness	
Schutz G-H, Assertiveness	.61
Schutz 5, Assertiveness	.66
Schutz 6, Success in Attaining Leadership	.57
Schutz 7, High Self-Estimate of Leadership Ability	.43
S Factor from Guilford's STDCR, Social Extroversion	.58
P Factor from Guilford's Opinion Inventory, High Level of Physical Drive	.36
Factor 2: Power Orientation	
Schutz A-B, Power Orientation	.43
G Factor from Guilford's Opinion Inventory, Self-Reliance	.40
Factor 3: Rigidity	
California F Scale, Authoritarianism	.56
S_o Factor from Guilford's Opinion Inventory, Tendency toward Orderliness	.39
F Factor from Guilford's Opinion Inventory, Cultural Conformity	.53
Factor 4: Aggressive Nonconformity	
V_e Factor from Guilford's Opinion Inventory, Need for Variety	.45
L Factor from Guildord's Opinion Inventory, Need for Attention	.54
Q Factor from Guilford's Opinion Inventory, Aggression	.62
M Factor from Guilford's Opinion Inventory, Resistance to Restriction	.43

Rotated factor loadings may be found in Table 1. The factors were interpreted as follows:

Factor 1, Assertiveness, was loaded heavily on Schutz's Assertiveness scores, as well as Guilford's Physical Drive score and Social Extraversion score. A separate Sociability Factor, mentioned as an important variable in several studies, did not emerge. If additional Sociability measures had been included in the predictor battery, the factor may have emerged or may have loaded under Assertiveness as did Guilford's Social Extraversion score.

Factor 2, Power Orientation, had loadings on Schutz's Power Orientation score, as well as Guilford's Self-Reliance variable.

Factor 3, Rigidity, seems clearly indicative of Rigidity, being loaded most heavily on the California F scale. Other significant loadings include Guilford's Cultural Conformity score and Orderliness score.

Factor 4, Aggressive Nonconformity, is more difficult to interpret. The heaviest loading on this factor was in Guilford's Aggression scores. Other variables loading on this factor included Guilford's Need for Variety score, Need for Attention score, and Resistance to Restriction score. Several interpretations of this factor were considered. In some respects it appeared quite similar to Cattell's Hypomanic Aggression. Another possible interpretation was that the factor was indicative of social immaturity, while a third interpretation was that it reflected frustration in the military environment. For several reasons, an interpretation of the factor as Aggressive Nonconformity seemed most plausible. First, the heavy weighting of Aggression and Need for Attention seemed to point in this direction. The loadings on Need for Variety and Resistance to Restriction suggest a negative reaction to the restrictions and conformity demands found in the military environment. Observations of the Command and Staff School, secondly, reveal that a significant number of students display overt aggressive attitudes toward the school. Some students also displayed considerable negativism concerning the testing itself. This negativism might well have been reflected in some of the test scores where it could have come to the surface. For example, some of the items on the Aggression test, such as those advocating punitive action, could give the *S* an indirect means of expressing aggression or negativism aroused by the testing situation. Weiss' (1956) research dealing with the effect of the individual's immediate mood on attitude measures indicates that the individual's mood may be reflected to a significant degree in items where aggression can be displayed.

Weighted factor scores were computed including all variables loading over .30 and most heavily loaded on the given factor. These personality factor scores constituted the predictors against which to compare small group role scores (see Table 1).

Small Group Role Scores

The small group role scores were based on two peer nominations completed in the second and fifth weeks after the groups were organized. The first peer nomination contained six brief paragraphs describing typical small group behavior patterns. These paragraphs described popular-social behavior, good followers, assertive behavior, rigid behavior, creative behavior, and leadership. Each seminar member nominated the person (by code number) who best fitted each description and checked a five-point scale to indicate the degree to which the selected individual fitted the description. Each seminar member received a score for each of the six roles which was the sum his nominations weighted for degree of fit. These scores were then converted to standard scores.

The second peer nomination used the same six role descriptions but required the individual to select three seminar members most fitting the description and three least fitting the description. The individuals' score for each role was the algebraic sum of his best fit and least fit nominations. Scores on the second peer nomination were converted to standard scores and combined with first peer nomination scores to give six composite role scores for each *S*.

RESULTS

In order to determine the degree to which the personality factor scores predicted small group role scores, samples of individuals who had made either high or low role scores were selected. The four individuals in each seminar who received the highest scores on a given role were selected as well as the four indi-

TABLE 2
CORRELATIONS BETWEEN PREDICTOR FACTORS AND SMALL GROUP ROLE SCORES

Factor Scores	Popular-Social 1	Good Follower 2	Assertive 3	Rigid 4	Creative 5	Leader 6
1—Assertive	.245*	−.221*a	.422*a	.216*a	.423*a	.389*a
2—Power Orientation	−.044	.066	.012a	.008a	.020a	.020
3—Rigidity	.065a	.026a	−.114*	.012	−.176*a	−.158*a
4—Aggressive Nonconformity	.108	−.060	.141*a	−.020	.207*a	.175

*Significant beyond .01 level.
aResults as hypothesized.

viduals with the lowest role scores. These cases were further screened to eliminate individuals from the top group who were not at least .6 σ above the mean on both peer nominations and eliminating individuals from the bottom group who were not at least .6 σ below the mean on both peer nominations. Thus, the persons chosen were not only at the extreme within their group but were at the extreme with respect to the entire sample on the role score under consideration.

A total of 24 specific hypotheses were then developed. The relationship of each of the four predictor factors to each of the six small group roles was hypothesized to be either positive, negative, or zero (see Table 2) based upon psychological theory and previous research results. Fourteen of these hypotheses were supported by the data. In all, 12 of the 24 widespread biserial correlations computed between the small group roles and the personality factors were significant at the .01 level.

It may be seen, however, in Table 2 that although many of the correlations were significant they were too small to permit prediction of the role of a given individual from his personality scores. By far the most promising of the predictor scores is Factor 1, Assertiveness. This score correlated significantly with all six role scores and yielded correlations around .40 with Assertiveness, Creativity, and Leadership roles. As Table 3 indicates, these three roles are moderately intercorrelated, suggesting that they are measuring a single broad leadership role. A combination of these three role scores correlates .46 with Factor 1. This correlation is sufficiently high to permit reasonably good selection for small group leadership if it is established that this role is stable from situation to situation.

TABLE 3
CORRELATIONS BETWEEN DIFFERENT PEER ROLES ON THE SECOND RATING

Variable	1	2	3	4	5	6
1. Popular-social	.000	.139	.255	−.348	.264	.252
2. Good Follower	.139	.000	−.251	−.404	.065	.105
3. Assertive	.255	−.251	.000	.452	.710	.599
4. Rigid	−.348	−.404	.452	.000	.207	.177
5. Creative	.264	.065	.710	.207	.000	.817
6. Leader	.252	.105	.599	.177	.817	.000

DISCUSSION

The results shown in Table 2 generally support the findings of earlier research. Several studies have been reported that are concerned with relationships between personality and small group behavior. Some of these have analyzed small group observational data so as to yield personality factors. Some have explored relationships between peer or supervisor evaluations and small group behavior. Only a few, however, have attempted to predict small group behavior from personality test scores.

Leadership ratings of 20 sorority girls in a leaderless group discussion situation were compared with their scores on the Rorschach, Guilford-Zimmerman Temperament Survey and F scale by Bass (1953) in one study of this type. Bass found no significant results with the F scale, but did find that Ascendence and Sociability scores on the Guilford-Zimmerman and High Verbal Output on the Rorschach were significantly correlated with Leadership ratings. The Rorschach result is in all likelihood an artifact of Bass' LGD rating system,

which depends to a considerable degree on Verbal Output (Bass, 1951). The correlation between Ascendency and Leadership in Bass's study, however, is almost identical to the correlation between Assertiveness and Leadership found in the research reported in this paper.

In another study of small group leadership (Cattell & Stice, 1953) leaders in 34 groups were compared in terms of scores on the 16 PF test. He found four variables that differentiated between leaders and nonleaders at the .05 level. These were Super Ego Strength, Adventurousness, Self Confidence, and High Self-Sentiment Formation. These variables, as described by Cattell, are similar in many respects to the Assertiveness and Power Orientation factors used as predictors in this paper. Richardson and Hanawalt (1944) compared two types of leaders and nonleaders on the basis of their scores on the Bernreuter Personality Inventory. They found their leader groups to be more dominant, extraverted, and self-sufficient than the nonleaders.

Thus, the few studies that have been done, despite the use of different personality measures, different leadership measures, and different types of *Ss*, all agree to a considerable extent that leaders are significantly more assertive, self confident, and extraverted than nonleaders.

Studies employing other personality variables as predictors of small group roles are much less conclusive.

Several workers have employed the F scale and related measures in small group research with varying results. The rigidity factor used in the research reported in this paper was correlated negatively to the variables in the broad leadership role, but the correlations, although significant, are too low to have much meaning. Some work has indicated that authoritarian leaders function effectively in authoritarian groups (McCurdy & Eber, 1954). There is also evidence that different types of individuals emerge as leaders in authoritarian and equalitarian groups (Haythorn et al., 1956; Adorno et al., 1950). Some work has demonstrated a positive relationship between Authoritarian Leadership and Group Cohesiveness (Medalia, 1955), while other research has found a negative relationship between these variables (Lippitt, 1940). Unlike Assertiveness, which appears to be an important leadership variable in most groups, the importance of authoritarianism varies greatly with the group, the task, and the leader.

Four of Schutz's items loaded heavily in the Assertiveness factor and one in the Power Orientation factor identified in this research. The results of Schutz's experimental battery, particularly the Assertiveness items, were impressive but indicated a need for further development. Schutz's scores have rather low reliability as they are each based on only one item. The items aimed at measuring Assertiveness were found to correlate with each other from .19 to .47. Items measuring Power vs. Personal Orientation, however, correlated with each other $-.09$ to $+.09$, indicating little common variance. As Schutz's technique becomes more highly developed, the Power Orientation variable might become useful in role prediction.

SUMMARY AND CONCLUSIONS

A sample of 819 Air Force officers was administered a test battery designed to predict the individual's role in small group situations. This battery was factor analyzed, yielding four factors: Assertiveness, Power Orientation, Rigidity, and Aggressive Nonconformity. The sample was then divided into 60 small groups of 12 to 14 members and rated with respect to six small group roles. Twenty-four widespread biserial correlations were computed between the four predictor factors and the six small group roles. Twelve of these correlations were significant at the .01 level. The predictor factor "Assertiveness" was most successful, correlating .46 with a composite leadership role.

With regard to significance of these results for future work, it seems reasonable to conclude from the success in predicting the leadership composite that prediction of certain roles and behavior patterns in small group activity can be achieved by further developing predictor instruments along the lines indicated by this study. Further research may make it possible to analyze changes in the individual's role behavior that are attributable to differences in certain measurable characteristics

of other group members in problem solving groups.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BASS, B. M. Situational Test: 2. Leaderless group discussion variables. *Educ. psychol. Measmt.*, 1951, **2**, 196–207.

BASS, B. M., MCGEHEE, C. R., HAWKINS, W. C., YOUNG, P. C., & GEBEL, A. S. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 120–128.

CARTER, L. F. Some research on leadership in small groups. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951.

CARTER, L. F. Evaluating the performance of individuals as members of small groups. *Personal Psychol.*, 1954, **7**, 477–484.

CARTER, L. F., HAYTHORN, W., & HOWELL, M. A further investigation of the criteria of leadership. *J. abnorm. soc. Psychol.*, 1950, **45**, 350–358.

CARTER, L. F., & NIXON, M. An investigation of the relationship between four criteria of leadership for three different tasks. *J. Psychol.*, 1949, **27**, 245–261.

CARTER, L. F., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.

CATTELL, R. B. *Handbook for the objective-analytic personality test batteries,* Institute for Personality and Ability Testing, Champaign, Illinois, 1955.

CATTELL, R. B. Determining syntality dimensions as a basis for morale and leadership measurement. In H. Guetzkow (Ed.), *Groups, leadership and men,* Pittsburgh: Carnegie Press, 1951.

CATTELL, R. B., & STICE, G. S. *Four formulae for selecting leaders on the basis of personality.* Urbana, Ill. Univ. of Illinois, Lab. Pers. Assessment & Group Behav., 1953.

GREER, F. L. *Small group effectiveness, Series 1955.* Philadelphia: Inst. for Res. in Hum. Relat., 1955.

GUILFORD, J. P. *Opinion Inventory.* San Antonio, Tex.: Air Res. and Devel. Command. 1951.

GUILFORD, J. P. *Manual. An inventory of factors STDCR.* Beverly Hills, Calif. Sheridan Supply, 1945.

HAYTHORN, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 265–284.

HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. The behavior of authoritarian and equalitarian personalities in groups. *Hum. Relat.*, 1956, **9**, 57–74.

LEWIN, K., LIPPITT, R., & WHITE, R. Patterns of aggressive behavior in experimentally created social "climates." *J. soc. Psychol.*, 1939, **10**, 271–299.

LIPPITT, R. An experimental study of authoritarian democratic group atmospheres. *Univ. Ia. Stud. Child Welf.*, 1940, **16**, 43–195.

MCCURDY, H. E., & EBER, H. W. Democratic vs. authoritarian: A further investigation of group problem solving. *J. Pers.*, 1954, **22**, 258–269.

MEDALIA, N. Z. Authoritarianism, leader acceptance, and group cohesion. *J. abnorm. soc. Psychol.*, 1955, **51**, 207–213.

RICHARDSON, HELEN M., & HANAWALT, N. G. Leadership as related to the Bernreuter Personality Measures: III. Leadership among adult men in vocational and social activities. *J. appl. Psychol.*, 1944, **28**, 308–317.

SAKODA, J. M. Factor analysis of OSS situational tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 843–852.

SCHUTZ, W. C. What makes groups productive? *Hum. Relat.*, 1955, **8**, 429–465.

WEISS, W., & FINE, B. J. The effect of induced aggressiveness on opinion change. *J. abnorm. soc. Psychol.*, 1956, **52**, 109–114.

(Received January 19, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CHARACTERISTICS AND CONCOMITANTS OF MOOD FLUCTUATION IN COLLEGE WOMEN

ALDEN E. WESSMAN, DAVID F. RICKS[1]
Harvard University
AND MARY McILVAINE TYL
Radcliffe College

THE pathologically extreme moods that occur in neurotic depressions and in the affective psychoses have been described in much the same way for centuries (Burton, 1621; Bellak, 1952); but the causes of these alterations are still debated, and their relations to normal variations in hedonic level are still unclear. The present study is based on the general proposition that the difference between pathological and normal mood changes is essentially one of degree rather than of type, and that the phenomena which have been so thoroughly described as characteristic of extreme moods also characterize, in less pronounced forms, the mood variations of normal individuals. This proposition is examined using three sets of data provided by a group of normal college women: (*a*) their self and ideal representations reported in relatively elated and relatively depressed states, (*b*) the direction of punitive responses and persistence of effort revealed by a projective frustration test taken in these states, and (*c*) the average level, the frequency, and the intensity of their mood fluctuations reported daily over a period of six weeks.

Few studies of hedonic level and affects in normal persons have been concerned with repeated measurements over a period of time. Flügel (1925), regarding previous studies as segmental, artificial, and of insufficient duration, reported an investigation of the temporal patterns and nature of affective processes of nine subjects who kept systematic and detailed introspective records of their feelings of "pleasure" and "unpleasure" and who made frequent self ratings on a seven-point hedonic scale for a minimum period of 30 days. He reported that the data for every subject showed a distinct preponderance of "pleasure" over "unpleasure," and that the duration of intense affective states (whether "pleasure" or "unpleasure") was considerably shorter than that of less intense feelings. He also found "evidences to suggest that those who tended to experience the most extreme degrees of feeling are on the whole less happy than those whose feelings are usually less intense."

Johnson (1937) asked 30 female college students to furnish daily self ratings on mood for 65 to 90 consecutive days and to serve in two experimental sessions, one in an elated and one in a depressed mood. She reported that the individual temporal patterns of daily mood showed varying degrees of rhythmic tendency. She found no apparent relation with menstrual periodicity. A fairly consistent feature of the temporal patterns was the longer duration of the elated as compared to the depressed moods. For individual subjects, measurable behavior differences were found in elated and depressed moods. The subjective correlates of depressed moods revealed, in lesser degree, the symptoms characteristic of affective psychopathology: reduced feelings of physical energy, felt loss of power and capacities, reduced outgoing friendliness, indecisiveness, diffidence, and withdrawal from social interaction.

The general proposition that differences between pathological and normal mood phenomena are primarily matters of degree rather than of type suggests a broad scale application of hypotheses derived from psychoanalytic investigations (Freud, 1917; Abraham: 1954a, 1954b; Rado, 1928; Fenichel, 1945) as well as the studies of normals reported above. The hypotheses below are of two varieties: first, hypotheses about general *intra-*

[1] This study was carried out by the senior authors in connection with research at the Psychological Clinic Annex, Laboratory of Social Relations, Harvard University, under Research Grants M-700 and M-1287 from the National Institute of Mental Health. (Project director: Henry A. Murray)

individual comparisons in elated and depressed states of (*a*) self and ideal conceptions, and (*b*) aggression and persistence of effort when frustrated; second, hypotheses relating *inter*-individual variations on the characteristic parameters of mood fluctuation to the relative stability and congruence of self and ideal conceptions.

Intra-individual Change in Depression and Elation

1. Self esteem, as measured by the correlation between the self and the ideal, is higher in elation than in depression.

2. The self concept changes more than the ideal concept with changes of mood.

3. The self concept is relatively favorable in elation, and relatively unfavorable in depression.

4. The ideal concept is relatively less "good" (and therefore less demanding) in elation, and relatively more "good" (and therefore more demanding) in depression.

5. There is greater internal (introceptive) preoccupation in depression than in elation, and less interest in the environment.

6. The frequency of intropunitive responses to frustration is higher in depression than in elation, and the frequency of extrapunitive responses is lower.

7. The frequencies of obstacle-dominated and ego-defensive responses to frustration are higher in depression, and the frequency of need-persistent responses is lower.

Concomitants of Interindividual Differences on the Parameters of Mood Fluctuation

8. Self consistency, as measured by the correlation between the self description in elation and self description in depression, is negatively related to (*a*) the relative intensity of mood fluctuations, and (*b*) the relative frequency of pronounced mood fluctuations.

9. Self esteem, as measured by the correlations between the self description and ideal description (in elation and depression) is positively related to average level of mood (i.e., general hedonic level).

METHOD

Subjects

Ss were 25 Radcliffe College students, all from the same dormitory, who volunteered to take part in a six-week study (October–December 1957) that required 5–10 minutes every night for recording their feelings, together with taking certain psychological tests. Of the original 25 *Ss*, 21 responded 30 days or more of the possible 42 days; their records were used in the final data analysis of the patterns of mood fluctuation. The tests of hypotheses, however, are based on the 14 *Ss* who provided a sufficient record of their feelings and who also completed the psychological tests described below.

Record of Daily Moods

Daily self ratings were made for six weeks on 11 scales describing different affective states. Each *S* also indicated daily the state of her health, the felt pressure of immediate academic work, and whether or not she was menstruating. Only the results of the 11th scale, *Mood*, and the question concerning menstruation are considered in this report.

The *Mood* scale contained 10 phrases, ranging from expressions of extreme depression through more neutral feelings to those expressive of extreme elation. The phrases were chosen so that (*a*) there would be approximately equal subjective gradations between units, and (*b*) the ends of the scale would be so extreme that few *Ss* would experience them, and these few would do so on rare occasions. This scale is given below:

10. Complete elation, rapturous joy, and soaring ecstasy.
9. Very elated and in very high spirits; tremendous delight and buoyancy.
8. Elated and in high spirits.
7. Feeling very good and cheerful.
6. Feeling pretty good, "OK"
5. Feeling a little bit low. Just so-so.
4. Spirits low and somewhat "blue."
3. Depressed and feeling very low. Definitely "blue."
2. Tremendously depressed. Feeling terrible, really miserable, "just awful."
1. Utter depression and gloom. Completely down. All is black and leaden. Wish it were all over.

The scale was marked each night just before retiring for: (*a*) the best *S* had felt during the day, (*b*) the worst *S* had felt during the day, and (*c*) *S*'s average for the day.

Assessment of Self Concept and Ideal Concept in Elation and Depression

A Q sort was constructed by asking each *S* to write 10 phrases which could describe her, a friend, or other girls at Radcliffe. Almost all of these descriptive phrases could be classified according to *area*, i.e., whether the phrase referred to (*a*) introceptive feelings, (*b*) interaction with peers, or (*c*) intellectual traits and work in college; and *evaluation*, i.e., whether it implied (*a*) favorable, (*b*) neutral, or (*c*) unfavorable judgment.

For the Q sort, 45 items were selected, 5 in each of the 9 categories of the design.[2]

When filling out the Q sort to describe self and ideal, the *S* was instructed to fit the 45 items to a forced distribution on a 7-point scale ranging from "−3, least characteristic," through "0," to "+3, most characteristic." The number of items to be fitted in the respective scale positions, −3 to +3, were 3, 5, 9, 11, 9, 5, 3.

This set of items was given to the *Ss* 3 weeks after the recording of their daily moods had begun, with the instructions that they were to wait until two occasions, one in which their mood reached an extreme "high," and one an extreme "low," and at those times administer the tests to themselves. The "elated" and the "depressed" sets were given to the *Ss* in sealed envelopes, and were identical, consisting of two sets of the same 45 items, the first (self concept) to be filled out "to give the most accurate picture of yourself as you really believe you are now," the second (ideal concept), to be done after the first and without comparing items, to be filled out "to give a picture of yourself as the kind of person you have hoped to become or have fancied yourself to be; the picture you describe should be the personal ideal for you." After the *S* filled in the sorts, she placed them in an envelope labeled with a pseudonym and gave them back to the experimenters. *Ss* were requested not to show the sorts to, or discuss them with, each other.

Projective Reactions to Frustration in Depression and in Elation

The Rosenzweig Picture-Frustration Test, Standard Revised Form for Adults, was used to measure projective reactions to frustration in depression and in elation (Rosenzweig, Fleming, & Clarke, 1947). Following some preliminary exploration, the instructions were changed slightly to read: "Imagine what *you* would say if you were the other person, and write in the box the very *first* reply that comes into your mind. Please avoid the temptation to be overly humorous. *Remember: we want what you probably would say in such a situation.*"

Two Rosenzweig P-F tests were given to each *S* with the Q sort forms and were self-administered in the same way—one when the *S* felt very elated and one when she felt very depressed.

The responses were scored according to Rosenzweig's instructions (*a*) as to *direction of aggression*, either "extrapunitive," "intropunitive," or "impunitive"; and (*b*) as to *type of reaction*, either obstacle dominance, ego defense, or need persistence; and the percentage of the total responses for each *S* in each of the classifications was obtained. The protocols were scored and hypotheses tested independently by two of the investigators. While there were some minor differences in scoring, the results of the statistical tests of the hypotheses were identical.

RESULTS

Intra-individual Change in Depression and Elation

Hypothesis 1. Self esteem. Of the 14 *Ss* completing this part of the experiment, 13 showed a higher correlation between the self and the ideal in elation than in depression ($p < .005$, sign test[3]). In depression the correlations ranged from .78 to −.83, with a median of .07; in elation they ranged from .87 to −.14, with a median of .55. The median difference between the elated self-ideal correlation and the depressed self-ideal correlation was .45. If correlation between self and ideal can be regarded as a measure of self esteem we can conclude that, for these *Ss*, one consistent concomitant of a change from elation to depression is a drop in self esteem.

Hypothesis 2. Relative Consistency of the Self and Ideal. The correlation between the ideal in depression and the ideal in elation was higher than the corresponding correlations for the self concepts for all 14 *Ss* ($p < .005$). The correlations between the two ideal sorts ranged from .93 to .72, with a median value of .84. The correlations between the self sorts ranged from .79 to .18, with a median value of .63. The median difference between the two correlations was .26. The data indicate that in mood swings, the ideal remains relatively more stable than the self concept.

Hypothesis 3. Favorableness of the Self Concept. For all 14 *Ss* the score for "unfavorable" statements regarding the self was greater, i.e., more characteristic, in depression than in elation, as predicted ($p < .005$). It was expected that this relation would hold for all 3 subareas of the Q sort; the relation was confirmed for the areas of "introception" (13 of 14 *Ss*, $p < .005$) and "interaction" (13 of 14 *Ss*, $p < .005$), but it was not supported for the area of "work" (8 of 14 *Ss*, not significant).

[2] The set of structured Q-sort items has been deposited with the American Documentation Institute. Order Document No. 6095, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] All of the probabilities reported in the section on intra-individual changes with mood are based on one-tailed sign tests (Mosteller & Bush, 1954, pp. 312–314).

For 13 of 14 *Ss* the score for "favorable" items regarding the self was smaller, i.e., less characteristic, in depression than in elation ($p < .005$). The predicted relationship held for all three areas of the Q sort: "introception" (12 of 14 *Ss*, $p < .025$), "interaction" (13 of 14 *Ss*, $p < .005$), and "work" (12 of 14 *Ss*, $p < .025$). The results bear out the hypothesis that depression is characterized by an increase in self-derogatory attitudes and a decrease in feelings of goodness and worth.

Hypothesis 4. Favorableness of the Ideal. It was predicted that the ideal concept, unlike the self concept, would appear more favorable, more "good," in depression than in elation (and thereby increase the distressing disparity between the ideally desired self and the presently perceived "real" self). This was not confirmed for either "positive" items (6 of 12 *Ss*, 2 ties, not significant) or for "negative" items (4 of 10 *Ss*, 4 ties, not significant).

Inspection of the three subareas in the Q sort showed that for "interaction" and for "work" there were no significant findings in either the predicted direction or its opposite. In the subarea of "introception," however, the ideal was significantly more "favorable" or "good" in depression (10 of 13 *Ss*, 1 tie, $p < .05$). For negative items the data approaches significance (8 of 11 *Ss*, 3 ties) in the predicted direction. We conclude that the ideal is relatively less affected by mood changes than is the self (Hypothesis 2); but that in the area of introceptive feelings and evaluations the ideal is more "good," and thereby more discrepant from the actual self, in depression than in elation.

Hypothesis 5. Degree of Introception. This hypothesis suggested a greater preoccupation with "introceptive" judgments in depression, with less emphasis upon the persons and pursuits involved in "interaction" and "work" items. The items were accordingly scored, not only for the degree to which they characterized the self, but also for their salience, as indicated by the magnitude of their positive or negative discrepancy from the mean. Although the trend was in accord with the prediction, the hypothesis was not confirmed. For several girls there appeared to be a movement from interest in interaction and introception in elation to preoccupation with work in depression.

Hypothesis 6. Direction of Punitive Responses. The results of the Rosenzweig P-F test did not support the hypothesis concerning the direction of punitive responses. Contrary to our prediction, the frequency of extrapunitive responses was significantly higher for this group in depression (11 of 14 *Ss*, $p < .05$). Intropunitive responses were significantly less frequent for the group in depression (11 of 14 *Ss*, $p < .05$). Impunitive responses tended to be less frequent in depression than in elation (9 of 13 *Ss*, 1 tie, not significant).

These results suggest that most *Ss*, in satisfied and elated moods, are less hostile towards external objects, but that in depression they have a tendency to project more extrapunitive responses.

Hypothesis 7. Persistence of Effort. As predicted, the frequencies of obstacle-dominated and ego-defensive responses to frustration elicited by the Rosenzweig P-F test were higher in depression, but not significantly so. The prediction that the frequency of need-persistent responses would be lower in depression than elation was significantly supported (11 of 14 *Ss*, $p < .05$). This set of findings is congruent with formulations such as Bull's which analyze depression as perseveration in helplessness and protest by a person who has become fixated on his own entanglements and discomforts in the face of frustrating interruptions of progress toward goals (Bull, 1951; Bull & Strongin, 1956).

Characteristics and Parameters of Mood Fluctuation

Mean Value of Average Mood. The individual means of the daily ratings of "average mood for the day" over the entire period varied from 5.43 to 7.37 for the 21 *Ss*. The mean of their means was 6.14, and the median was also 6.14. The midpoint of the mood scale is 5.5, half way between "5. Feeling a little bit low, just so-so," and "6. Feeling pretty good, OK." With but one exception, the mean of the average mood of each *S* fell above 5.5;

in other words they tended more toward "elated" than "depressed" moods. The data are consistent with a substantial body of findings regarding avowals of "average level of happiness" (Goldings, 1954; Wessman, 1956).

Intensity of Mood Fluctuation. Two possible measures were considered for ascertaining intensity of mood swings. One was the variability of the "daily average" mood, as measured by its standard deviation, which indicated the degree of fluctuation of the average daily mood from its mean. This measure ranged from .42 to 1.81 in the group, with a median of .94. The other measure considered was the average range of feeling between the "lowest mood" and "highest mood" for each *S* for each day; thus it represents a measure based on the extremes of feeling of the *S* for each day. This "average daily range" varied from 1.17 to 4.36 for the group, with a median of 2.61.

To ascertain whether these two measures were comparable, whether those *S*s whose average mood fluctuated a great deal from day to day were also the *S*s given to the widest ranges of mood alteration within single days, the two were rank ordered and correlated. The result, though not statistically significant, may indicate some degree of positive correlation ($r = .34$). As both measures of intensity are of interest, both were used and are reported in the data analysis.

Our data did not confirm Flügel's (1925) inference that those who "experience the most extreme degrees of feeling are on the whole less happy than those whose feelings are usually less intense." The rank order correlation of the mean value of the average mood with the standard deviation was $-.17$, and with average daily range was $+.36$, neither of which is significant.

To see if there was a relation between the tendency to have high extremes and the tendency to have low extremes *S*s were ranked on (*a*) the average of the 5 "lowest" scores on the "lowest moods" and (*b*) the average of the 5 "highest" scores on the "highest moods." The rank orders correlated .49 ($p < .025$), indicating the *S*s who reported the most depressed moods also tended to be those who reported the most elated moods.

Frequency of Pronounced Mood Fluctuations. Alternative ways of measuring frequency of fluctuation were considered. A count of the number of times an *S*'s "daily average mood" crossed its mean was considered and rejected because it would make no distinction between relatively minor fluctuations about the mean and more pronounced alterations of mood—and it is usually the latter which is referred to when characterizing an individual as "moody" or "changeable." The measure of frequency adopted instead was a count of the number of periods that an *S*'s "average daily mood" was outside of the values 5, 6, and 7 on the mood scale (the values within which most of the fluctuations occurred). Thereby every *period* (one day or several days) the *S*'s mood was "8. Elated and in high spirits" or more extreme on the elated side, or "4. Spirits low and somewhat 'blue'" or more extreme on the depressed side, was counted as one. Thus this was a measure of the *frequency of pronounced mood swings.* Comparing these scores with visual inspection of the *S*'s records indicated it to be a satisfactory and representative measure of the phenomena. The *S*s ranged from 1 to 12 pronounced peaks and/or troughs over the 42-day period.

Patterning of Mood Fluctuation. Inspecting the plots of the *S*s' daily reports, it is apparent, as it was in Johnson's study (1937), that the fluctuations have more of an *m* shape than a *w* shape. That is, the periods of depression are generally represented by a sharp dip in the graph, usually of short duration, often only a day; the periods of relative elation continue over longer periods of time. This finding corresponds with the finding that the *S*s report feeling, on the whole, more often relatively elated than depressed.

Contrary to Johnson's interpretation of her observations, we did not observe any particular rhythm to our *S*s' fluctuations. Rather the periodicity and duration of fluctuations was decidedly irregular for all but one of the *S*s.

Relation of Menstruation and Mood. Johnson's study (1937) reported finding no apparent relationships of mood fluctuation with

menstrual periodicity. This contradicts both folklore and investigations like Benedek and Rubenstein's (1942) which found feelings of frustration and irritability more prominent a few days before the menstrual period. Our data showed 14 out of the 19 *Ss* reporting menstrual periods had an "average" mood lower than their general mean during the 2 days prior to menstruation ($p < .05$), and 8 of the 19 reached their lowest moods of the entire 6-week study on these 2 days. Of the 19 *Ss*, 11 also showed an "average" mood lower than their mean on the day the menstrual period commenced.

Concomitants of Interindividual Differences on the Parameters of Mood Fluctuation

Hypothesis 8. Intensity and Frequency of Hedonic Fluctuation and Relative Consistency of the Self Concept. The correlations between the self descriptions provided in depression and in elation, which ranged from .18 to .79, with a median of .63, provided the basis for a rank order of the 14 *Ss* on relative consistency of the self concept. Rank order correlations were then computed against ranks on intensity of fluctuation and on frequency of pronounced mood fluctuations. Both parts of the hypothesis were significantly confirmed. Self consistency was negatively related to both measures of intensity of fluctuation, the standard deviation of the daily average ($r = -.64$, $p < .01$) and the average daily range ($r = -.46$, $p < .05$). Self consistency was also negatively related to frequency of pronounced mood fluctuations ($r = -.63$, $p < .01$). These data support the hypothesis that girls who have relatively intense mood changes, or relatively frequent ones, have less stable conceptions of their own identity. Average mood level, on the other hand, was unrelated to self consistency ($r = .09$).

It was shown above (Hypothesis 2) that the ideal concept changed less than the self concept with changes in mood. The small changes which did occur in the ideals were not significantly related to the intensity, frequency, or average level of the *Ss*' mood fluctuations.

Hypothesis 9. Average Hedonic Level and Relative Self Esteem. The measure of self esteem was the one used in Hypothesis 1, the correlation between the self and ideal descriptions. The *Ss* were rank ordered on these correlations in both elation and depression, and these ranks were correlated against a rank order on the mean "average daily mood" over the six-week testing period. Contrary to prediction, the correlations were negative (in elation $r = -.36$, not significant; in depression $r = -.50$, $p < .05$).

No significant relationships between either intensity or frequency of mood fluctuation and relative degree of self esteem were found for either elated or depressed states. Rank on self esteem in depression was only moderately correlated with self esteem in elation ($r = .40$, not significant).

The high negative correlations found between self esteem (correlation between self and ideal Q sorts) and happiness (mean "average daily mood") required further exploration. Two explanations seemed possible: 1. The happiness score might be in error, or reflect some quality other than happiness. Error seemed unlikely, since the scale was based on a six-week average of the "average daily mood" score and was therefore unlikely to have been greatly influenced by transient moods, misunderstanding of the scale, etc., and because an analysis of variance had shown that the between-subject variance, as compared to that within the individual daily records, was significant well beyond the .01 level. It is possible, however, that the average position on the happiness scale might be influenced by characteristics other than happiness, e.g., response sets to score high or low, or a set in some *Ss* to deny unhappiness. 2. The self-ideal correlation might measure qualities other than self esteem. A high correlation between the self and the ideal might reflect denial of undesirable qualities in the self (Block & Thomas, 1955), a form of ego restriction which might be more typical of unhappy people than of happy ones. Or the same high correlation might reflect restriction of aspirations, disgruntled acceptance of limitations, and lowering of ideals.

The content of the Q-sort self and ideal descriptions was examined for evidence bearing on the possibilities outlined above. The group of *Ss* was divided in half according to their

mean "daily average mood" over the six-week observation period, and the Q-sort descriptions provided by the seven relatively more happy girls were compared with the descriptions provided by the seven relatively less happy girls. The typical self and ideal characterizations of the "happy" and the "unhappy" girls are given below. Then the significant discrepancies between the self and ideal descriptions are summarized and discussed, first for the happy and unhappy girls in elation, then for both groups in their relatively depressed moods.

Self Descriptions. The composite self picture presented by the happy girls, both in depression and in elation, was that of a friendly, altruistic person with diverse interests and some sense of perspective about herself, able to enjoy work without being preoccupied with it, and not considering herself sophisticated or poised. The change from elated to depressed moods seemed to be characterized by a decrease in breadth of interests and in satisfaction with learning, and an increase in nervousness, worried concern over school work, introspection, and concern with emotional security.

The unhappy girls described themselves, in both moods, as introspective, altruistic, idealistic, sympathetic and understanding of the problems of others, and concerned with academic work.

Comparison of the self descriptions of the two groups indicated that the unhappy girls described themselves as more introspective, independent, sophisticated, self-sufficient, critical, and more interested in academic work; the happier girls described themselves as more friendly, sociable, unorganized, nervous, tolerant, more willing to give of themselves to others, and more interested in what others had to offer.

Ideal Descriptions. In their elated moods both groups had ideals which were alike in most respects, emphasizing qualities of being interesting and stimulating, having a good sense of humor about themselves, and being sensitive and perceptive of the feelings of others. The happy girls tended to place a higher value on friendliness and consideration for others, while the unhappy girls placed a higher value on work.

In depression both groups valued work more than in elation, but this emphasis on work was much more extreme for the unhappy girls. For the happy girls concern with work in depression took the form of putting a higher value on the inherent satisfactions in learning and greater concern with discovering what they wanted out of life; for the unhappy girls it took the form of more exclusive concern with intellectual creativity and achievements.

The data indicate that one of the suggested explanations for the high self-ideal correlations of the unhappy girls, that they had lowered their ideals, is not valid. It appears more probable that the unhappy girls had *narrowed* their ideals, particularly those ideals they described in moods of depression, to exclude social values and place a more exclusive emphasis on scholastic and intellectual achievement.

Discrepancies between Self and Ideal. Examination of those items for which there were significantly consistent discrepancies (sign test) between selves and ideals showed 10 such items for the happy and 9 for the unhappy girls in depressed moods, 5 for the happy and 0 for the unhappy in elated moods. These differences indicate that in depression both groups experienced discrepancies between what they were and what they would like to be but that the unhappy girls, as a group, felt no consistent divergence from their ideals in elation, while the happy girls continued to feel areas of discontent. For the happy girls most of the discrepancies between actuality and aspiration concerned social and emotional qualities: ability to be interesting and stimulating, poise and self-confidence, and emotional security. The unhappy girls, in contrast, experienced discrepancies primarily in terms of unfulfilled intellectual ambitions. Elation seemed to be experienced by the unhappy girls as satisfaction of these internal standards, while for the happy girls elation was accompanied by outgoing interest in others and desire for the qualities which would enable this interest to be realized.

DISCUSSION

Intra-individual variation in psychological measurements has generally been ignored or regarded as a statistical inconvenience; a comparatively small number of investigations has made such variations a subject of study (Fiske & Rice, 1955) and only a few psychologists have stressed the value of methods of repeated measurement, or systematically employed them (Horn, 1950; Cattell, 1957). It is clear that repeated measurements play a basic role in other fields, particularly physiology and economics, and it is equally clear that some psychological characteristics, such as moods, vary markedly within individuals as well as between them. One aim of this study was the development of techniques for repeated measurements of moods and related feelings. The "Mood scale" used in this investigation has not yet been shown to have such desirable scale attributes as unidimensionality, equality of intervals, interindividual comparability, etc. For the present we can report that the scale seemed clear and relevant to the *S*s, and that their changes on the scale were related to demonstrable differences in self conception and projective reactions to frustration. Further study of this scale, and of the other scales whose details are not considered here, is currently in progress.

It should be recognized that the moods and self conceptions reported here are conscious avowals of subjective feelings and beliefs, and their validity is limited to an unknown extent by possible deficiencies in some *S*s' self awareness and capacity to experience, and perhaps by unwillingness or inability to report subjective states accurately. The data show only some aspects of the phenomenology of moods, and only by inference can this material be related to the unconscious dynamics central to affective life.

Although 25 *S*s commenced the study, the tests of the hypotheses are based only on the 14 *S*s who provided sufficiently complete data; there may be some systematic bias introduced by the omission of those who did not provide complete records. The use of *S*s from a homogeneous but relatively unstudied environment probably reduced extraneous and irrelevant variability within the sample to a near minimum; but it also made uncertain the degree to which some of the relationships found might reflect particular factors in the Harvard-Radcliffe subculture. The differences we found between the relatively happy and relatively unhappy girls may not indicate consistent differences between happy and unhappy people generally but, rather, reflect the particular role conflicts and strains at Radcliffe or similar women's colleges. It appeared that in this group the less happy girls had identified more with a "student" or "scholar" role, the more happy girls with a "date" or "prospective wife" role.

The major concern of the study was to test the general hypothesis of continuity and similarity of dynamic patterns between normal and pathological moods. Pathological moods have been more thoroughly studied and reported than normal moods, and there is substantial agreement as to their characteristics. A fairly broad and representative set of generalizations could therefore be drawn from the literature dealing with affective psychopathology and stated as hypotheses. Most of the findings which compared elated and depressed states in the same girls (i.e., intra-individual changes) supported the continuity hypothesis. Self esteem, as measured by the self-ideal correlation, was shown to be lower in depression than in elation, and the change was shown to be due mainly to a drop in feelings of goodness and worth in the self, with a possible secondary cause being an increase in the stereotyped "goodness," and presumably demanding character, of the ideal. Responses to the Rosenzweig P-F test indicated, as predicted, that the girls were less likely to go ahead to overcome obstacles in depression than in elation. One deduction from the continuity hypothesis was significantly reversed: *S*s were not, as had been predicted, more intropunitive in depression (Rosenzweig P-F test); rather they tended to be more punitive generally and were significantly more extrapunitive. There is general consensus in the clinical literature that anger and aggression are predominantly turned inward in severe depression. Our results suggest that the point at which mounting hostility

shifts from an outward to an inward direction may mark the transition from relatively mild depression to severe affective pathology.

Comparison of the *S*s with each other (interindividual comparison) on major parameters of mood change indicated that girls with stable self conceptions were less prone to experience either pronounced or frequent mood fluctuations. Priority of cause and effect relationships cannot be established here; probably individuals who have experienced marked and frequent changes in mood, either from physiological causes or due to developmental traumas, do not develop stable identities, and in later life the lack of a stable identity makes then more liable to such emotional storms.

A third interindividual comparison found that individuals with high self esteem (as measured by self-ideal correlation) were relatively unhappy, while individuals with low self esteem were relatively happy. Initially it may seem that this finding contradicts the earlier finding that *S*s had higher levels of self esteem when they were elated and lower levels of self esteem when they were depressed. The earlier finding, however, was based on an intra-individual comparison (each girl in elation compared with herself in depression), while the latter finding is based on an interindividual comparison of girls who were characteristically more happy with other girls who were characteristically somewhat less happy. It is logically fallacious to assume that relationships which can be demonstrated in intra-individual comparisons necessarily hold in interindividual comparisons, and vice versa; but often the experimental convenience of interindividual comparisons (all *S*s tested at one sitting, etc.) causes psychologists to overlook this limitation, and to argue that group comparisons support or contradict hypotheses which could be directly tested only by intra-individual techniques.

Even taking into account the logical differences between inter- and intra-individual tests of hypotheses, the finding of a negative correlation between self esteem and hedonic level still seemed paradoxical and required exploration of the content of the self and ideal descriptions as related to characteristic hedonic level. When the self and ideal descriptions of the more happy and less happy girls were compared they were found to show two contrasting but internally consistent patterns of feelings, interests, and aspirations. The unhappy girls resembled Hamilton and Mann's description of the predepressive personalities of patients with involutional psychosis (melancholia) who were found "to take life very seriously, to follow rigid patterns of behavior, to display lack of imagination and a sense of humor," and to be unsociable, lack spontaneity, have patterns of careful intellectual control, and to command respect "because of their thoroughness, conscientiousness, meticulous devotion to duty, and above-average level of productiveness." (Hoch & Zubin, 1954, p. 201). While it would be ridiculous to regard half of the *S*s as predepressive, this close correspondence does suggest that there is considerable resemblance between the personalities of the girls whose average level of happiness is below the median in this group and the personalities of those who eventually manifest depressions of pathological depth.

SUMMARY

A set of 10-point self-rating affective scales was developed for repeated daily measurements over extended periods. Records were kept over a 42-day period by 14 female college students who also took two identical sets of psychological tests (self and ideal Q-sort descriptions, Rosenzweig P-F test), once when they were elated, once when they were depressed. Hypotheses were tested as to intra-individual changes in elation and depression and as to concomitants of interindividual differences on the parameters of mood fluctuation. The significant intra-individual changes in depression (relative to elation) were: lower correlation of self and ideal, more unfavorable self concept, increased *extra*punitive responses; and reduced need-persistive responses. Individual records of mood fluctuation were markedly irregular and showed no clear periodicity. Generally the average mood was slightly on the elated side, and depressed periods were of short duration. The days prior to onset of menstruation were generally depressed. Relative interindividual standings on the parameters of mood fluctuations showed

the following significant correlations: intensity of mood swings inversely related to relative stability of the self concept; frequency of pronounced mood swings inversely related to relative stability of the self concept; and average mood (or general hedonic level) inversely related to correspondence of self and ideal. The last finding, which was contrary to prediction, led to an examination of the contents and discrepancies between self and ideal descriptions in elation and depression for the relatively more and relatively less happy girls. The less happy girls appeared to put more stress on intellectual achievements and their periods of elation seemed to correspond with meeting their high standards, while the happier girls appeared to have more diverse interests and be more sociable and even in elation experienced discrepancy between self and ideal. The relevance of these results to a general hypothesis of continuity between normal and pathological functioning was considered. The results suggest the value for personality research of systematic measurement of emotional fluctuations over extended periods, and of the consideration of affective states in personality testing.

REFERENCES

ABRAHAM, K. Notes on the psychoanalytical investigation and treatment of manic-depressive insanity and allied conditions (1911). In *Selected papers of Karl Abraham*. New York: Basic Books, 1954. (a)

ABRAHAM, K. A short study of the development of the libido, viewed in the light of mental disorders (1924). In *Selected papers of Karl Abraham*. New York: Basic Books, 1954. (b)

BELLAK, L. *Manic-depressive psychosis and allied disorders*. New York: Grune & Stratton, 1952.

BENEDEK, T., & RUBENSTEIN, B. B. The sexual cycle in women, the relation between ovarian function and psychodynamic processes. *Psychosom. Med. Monogr.*, 1942, **3**, Nos. 1 and 2.

BLOCK, J., & THOMAS, H. Is satisfaction with the self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254–259.

BULL, NINA. The attitude theory of emotion. *Nerv. ment. Dis. Monogr.*, 1951, No. 81.

BULL, NINA, & STRONGIN, E. The complex of frustration: A new interpretation. *J. nerv. ment. Dis.*, 1956, **123**, 531–535.

BURTON, R. *The anatomy of melancholy*. 1621.

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers-on-Hudson: World Book, 1957.

FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.

FISKE, D. W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.*, 1955, **52**, 217–250.

FLÜGEL, J. C. A quantitative study of feeling and emotion in everyday life. *Brit. J. Psychol.*, 1925, **15**, 318–355.

FREUD, S. Mourning and melancholia. (Originally published 1917). In *Collected papers*. Vol. 4. London: Hogarth, 1953.

GOLDINGS, H. J. On the avowal and projection of happiness. *J. Pers.*, 1954, **23**, 30–37.

HOCH, P. H., & ZUBIN, J. (Eds.) *Depression*. New York: Grune & Stratton, 1954.

HORN, D. Intra-individual variation in the study of personality. *J. clin. Psychol.*, 1950, **6**, 43–47.

JOHNSON, WINIFRED B. Euphoric and depressed moods in normal subjects. *Character & Pers.*, 1937, **6**, 79–98, 188–202.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Cambridge, Mass. Addison-Wesley, 1954. Pp. 289–334.

RADO, S. The problem of melancholia. *Int. J. Psychoanal.*, 1928, **9**, 420–438.

ROSENZWEIG, S., FLEMING, EDITH E., & CLARKE, HELEN J. Revised scoring manual for the Rosenzweig Picture-Frustration study. *J. Psychol.*, 1947, **24**, 165–208.

WESSMAN, A. E. A psychological inquiry into satisfactions and happiness. Unpublished doctoral dissertation, Princeton Univer., 1956. (Mic. 57-2228, University Microfilms, Ann Arbor, Mich.)

(Received February 5, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

REINFORCEMENT OF VERBAL AFFECT RESPONSES OF NORMAL SUBJECTS DURING THE INTERVIEW[1]

KURT SALZINGER AND STEPHANIE PISONI

New York State Department of Mental Hygiene and Biometrics Research, Columbia University

THE recent application of behavior theory to the study of verbal behavior has stimulated a great deal of research as is well evidenced in two recent reviews of the area (Krasner, 1958; Salzinger, K., 1959). The extension of this approach to the experimental analysis of the interview has been conceptualized by the senior author in a recent paper (K. Salzinger, mimeo edition, 1959) and has met with some success in actual experiment. So far, it has been shown that the output of verbal affect responses of schizophrenics can be manipulated by the use of discriminative stimuli such as questions (Suzanne Salzinger, 1956) and by the use of reinforcements like verbal expressions of agreement (Suzanne Salzinger, 1956; K. Salzinger & Pisoni, 1957, 1958a, 1958b). In addition, it has been shown that the effectiveness of conditioning is independent of the placement of the conditioning period in the interview, i.e., at the beginning or in the middle (K. Salzinger & Pisoni, 1957). Finally, the rate of extinction of affect responses has been found to vary directly as a function of the rate of acquisition (K. Salzinger & Pisoni, 1958b).

The present study attempts to extend the above findings by conditioning affect statements in a sample of normal individuals and comparing the results obtained with those of a sample of schizophrenics.

METHOD

Subjects

Eighteen female and eight male normal *S*s ranging in age from 20 to 50 years with a median age of 37.5 were selected from among the admissions to a general medical hospital. Patients with obvious psychosomatic symptoms such as ulcers and hypertension and those who showed evidence of any history or gross current signs of psychosis were excluded from the sample. Included were patients with illnesses such as pneumonia, tumors, nephrosis, gall bladder disorders, yellow jaundice, arthritis, and appendicitis, and those requiring minor surgical operations.

Experimental Procedure

Each *S* was interviewed for a period of 30 minutes. Thirteen *S*s were interviewed by a female *E* and 13 by a male *E*. All interviews were recorded with the recording equipment placed in full view of the patient. *S*s were informed that they had been selected randomly and that these interviews were being conducted in order to determine how such an interview might be incorporated into the routine medical examination administered in the hospital.

The interview began with specific questions concerning the name of the patient, his age, marital status, number of children and number of siblings. All the patient's responses to these questions were reinforced by *E*'s repetition of the response given, by his writing it down, and by his emission of phrases such as "mmm-hum," "uhhuh," "I see," etc. This procedure was adopted in order to obtain factual information upon which subsequent interview questions could be based and to establish the interviewer as a source of reinforcement, in this way encouraging the patient to speak in the presence of *E*. Following this, the experimental part of the interview was initiated with the question: "Would you tell me why you are here in this hospital?" Subsequently, the interviewer asked additional general questions whenever the patient ceased talking for more than two seconds.

Each interview was divided into three 10-min. periods. During the first 10 min. (operant level), the base rate of affect responses was determined. *E* asked questions but did not reinforce any statement made by the patient. Reinforcement was defined as verbal agreement consisting of such phrases as "mmm-hum," "uhhuh," "I see," and "yeah," immediately following statements made by the patient. During the second 10 min. (conditioning), *E* continued to question the patient and, in addition, reinforced each affect response by immediately following it with verbal agreement. During the last 10 min. (extinction), *E* withheld all reinforcement but continued asking questions whenever they were necessary to maintain the flow of speech of the patient.

Definition of Response

The response class of affect used in this experiment was defined as any statement by the patient describing

[1] This study was supported by Research Grants M586 (C3) and M1541 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The authors are indebted to J. Zubin and S. Sutton for their constant help and interest in this study. The authors are also grateful to L. C. Kolb and R. Senescu, who made the normal population available for this study, and to B. Schoenberg, who very generously gave of his time in the selection of *S*s. This paper was presented in part at a meeting of the Eastern Psychological Association, Philadelphia, Pa., 1958.

or evaluating his own state (other than intellectual or physiological). The response class therefore included only those affect statements which begin with the pronouns "I" or "we" as the following examples indicate: I am satisfied, I'm happy, we enjoyed it, I like him, I'm very close to him, I was mad at him, we hated him, I'll always be jealous of him, I am upset, I am a lonely person, I was so ashamed, I'm sorry for him, I feel (followed by any words), I was frightened, we couldn't take it, I always suffered. Quotations in which affect was ascribed to the speaker by another individual (e.g., "My husband said I didn't feel good"), although fitting all other criteria, were excluded on the basis of not being direct expressions of the *patient's* affect.

Statements like "I am happy and excited" were considered as one affect statement only, because the pronouns "I" or "we" did not preclde the second affective word. On the other hand, incomplete statements (in the sense that the object of the affect is not mentioned) like "I love..." or "we feared..." were viewed as bona fide responses.

Certain types of private events or internal states were excluded from the response class of affect because they referred primarily to intellectual processes, to actions which are sometimes but not always associated with affect, or to desires which appear to constitute a class of responses different from the affect class as defined here. Some examples are the following: I am confused, I am confident, I would like to..., I want, I was surprised, I am not well, we forgave him, I threaten her constantly, I didn't trust them.

The reliability of coding by the interviewers was examined in an earlier study (K. Salzinger & Pisoni, 1958a), yielding proportions of agreement ranging from .79 to 1.00, showing that both *E*s would generally have reinforced the same responses in identical interviews. Comparable coefficients of agreement were obtained for the interviews of this study.

A count per minute was taken of statements belonging to the general class of self-references (statements beginning with "I" or "we" which included among them self-referred affect statements) in order to compare changes in the occurrence of this class with those of the class of self-referred affect statements.

Interviewer Questions

Some or all of the following topics were discussed during each interview: reasons for being in the hospital and causes for illness; patient's relationships to his parents, siblings, fellow employees, employers, fellow students and teachers, wife or husband, children, friends; patient's activities during free time, and plans for the future. *E* made an attempt to balance these topics over the three conditions. For instance, if the patient discussed the symptoms of his illness in the operant level condition, *E* asked questions regarding the possible causes of the illness during conditioning and brought the patient back to these topics during extinction. As long as the topics were approximately balanced over the three conditions, however, *E* took his cues for topics from the content of the patient's statements. Both the number of topics and the order in which they were discussed varied from interview to interview.

Questions designed for eliciting affect directly, such as, "How did you feel about that?" or "Were you happy?" were not used.

RESULTS

Conditioning Effect for the Entire Group

The difference in number of affect statements emitted during operant level, conditioning, and extinction was tested by Friedman's nonparametric analysis of variance (Siegel, 1956) and found to be statistically significant ($p = .05$). The greatest number of affect statements was found during conditioning (sum of ranks = 60.5), the next greatest during extinction (sum of ranks = 52.5), and the smallest during operant level (sum of ranks = 43.0).

When the same statistical procedure was applied to general self-referred statements, no statistically significant difference was found ($p = .72$) among the three periods. Examination of the trends revealed that the largest number of self-referred statements occurred during operant level (sum of ranks = 55.0), the next largest during extinction (sum of ranks = 51.5), and the smallest during conditioning (sum of ranks = 49.5).

Conditioning Effect for Subgroups

In previous studies (K. Salzinger & Pisoni, 1957; 1958b), it was demonstrated that the conditioning effect takes place if *S* is given more than a certain minimum number of reinforcements. *S*s who received less than the minimum evidenced no changes in behavior as a result of the delivery of reinforcement. This finding held when a given number of reinforcements was assigned to each *S* before the interview as well as when reinforcements were made available only for a given length of time. When the present group of normals likewise was divided into those *S*s who received less than 8 reinforcements and those who received 8 or more, a significant difference in the number of affect statements between the conditions of operant level, conditioning, and extinction was found for the high reinforcement group ($p < .01$) but not for the low reinforcement group ($p = .90$). The high reinforcement group revealed the greatest number of affect statements during conditioning (sum of ranks = 32.5), the next

greatest during extinction (sum of ranks = 28.0), and the smallest during operant level (sum of ranks = 17.5). Examination of the general class of self-referred statements for the high and low reinforcement groups indicated no statistically significant difference among the three periods of the interview; the p level for the high reinforcement group was .90, for the low reinforcement group, .64.

Comparison of Normal and Schizophrenic Subjects

The normal *Ss* of this sample were compared to a sample of schizophrenic patients from a previous study (K. Salzinger & Pisoni, 1958a). These schizophrenics had been subjected to the same procedure as were the present *Ss*. Comparisons were made separately on each condition by the Mann-Whitney U test (Siegel, 1958). No statistically significant difference was found. The p value for the difference between operant levels was .20, for conditioning, .17, and for extinction, .87.

Members of the normal group were then matched with members of the schizophrenic group in terms of the following variables: sex, operant level, and number of reinforcements received. A pair was considered matched when the members were of the same sex, when the number of reinforcements received by each member differed by no more than 2, and when the discrepancy between the operant levels was no greater than three affect statements. Twelve pairs of individuals were found to meet these requirements. Although *Ss* were not matched on rate of acquisition, application of the paired replicates test (Siegel, 1956) to the normal–schizophrenic pairs revealed no significant difference ($p = .10$) with respect to this variable. The matched normals and schizophrenics were then compared on their resistance to extinction.

Application of the paired replicates test to these pairs revealed greater resistance to extinction in normals than in schizophrenics ($.02 < p < .05$). Nine of the normals extinguished more slowly than the schizophrenics; two showed no difference, and one schizophrenic extinguished more slowly than his normal match. The median difference in number of extinction responses was 6, the normals emitting 6 more affect responses, on the average, than did the schizophrenics during extinction.

The same comparison of individuals matched on operant level alone or on number of reinforcements alone did not serve to differentiate normals from schizophrenics ($p > .10$). Figure 1 represents individual cumulative curves of affect responses for two pairs of matched schizophrenics and normals over the three conditions. One matched pair has a low operant level, the other a high operant level. Inspection of this figure reveals both the similarity of the operant level and conditioning rates for the matched pairs as well as the difference in their extinction rates.

DISCUSSION

The results of this study resemble those previously obtained (K. Salzinger & Pisoni, 1958a). The interviewer's verbal reinforcement increased the frequency of uttering affect statements. Data taken from subgroups make it clear that a certain minimum number of reinforcements is necessary for conditioning to take place; i.e., an *S* who is given less than this number generally shows no evidence of conditioning. Investigation of the frequency of general self-referred statements showed it to be invariant over the three conditions of the interview, thus yielding evidence for the specificity of the effect of reinforcement upon self-referred affect statements.

While the finding that normals show greater

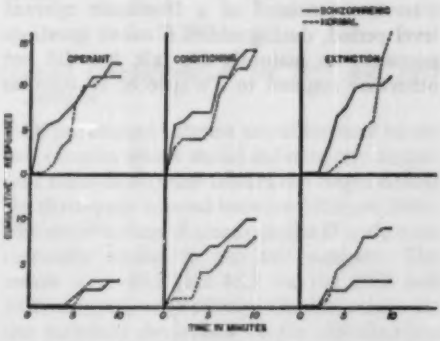


FIG. 1. Individual cumulative affect response curves for two pairs of matched schizophrenics and normals. The top pair shows two subjects with high operant levels, while the bottom pair shows two subjects with low operant levels.

resistance to extinction than schizophrenics is in agreement with the clinical observation of greater affect in normals, it suggests that the clinically "recognized" differences in amount of affect expressed are primarily a function of *S*s' differential reaction to the cessation of reinforcement. It now becomes necessary to see whether this difference in extinction holds only for the response class of affect or whether it applies to any response class. Such a study is presently in progress.

The difference in extinction rate has an important implication for psychotherapy with schizophrenics if the application of reinforcement is a useful tool in therapy. Some evidence for this contention has been made available by Murray (1956). The implication is that even though schizophrenics may appear to profit from psychotherapy (administration of reinforcement may result in conditioning), the profit (conditioning effect) may not persist as long in such patients as it does in normals after the therapy (reinforcement) is terminated. It remains for further studies to determine how much larger the number of verbal reinforcements must be for schizophrenics than for normals in order for the former to gain as much therapeutic profit from reinforcement as the latter.

SUMMARY

Twenty-six normal individuals, hospitalized for a variety of physical complaints, were interviewed for a period of 30 minutes. Each interview consisted of a 10-minute operant level period, during which *E* asked questions necessary to maintain *S*'s talk but did not otherwise respond to *S*'s speech; 10 minutes of conditioning, during which *E* reinforced by agreement all self-referred affect statements; and 10 minutes of extinction, during which *E* withheld all reinforcement. Conditioning of the verbal response class of self-referred affect statements occurred, and normal *S*s showed greater resistance to extinction than did schizophrenics.

REFERENCES

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MURRAY, E. J. A content-analysis method for studying psychotherapy. *Psychol. Monogr.*, 1956, **70** (13, Whole No. 420).

SALZINGER, K. The experimental approach to the interview. In J. Zubin (Ed.), *Experimental abnormal psychology*. Available in mimeo form at Columbia Univer. Library, 1959.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement verbal affect responses of schizophrenics during the clinical interview: The effect on conditioning of placement of the period of reinforcement. Paper read at the American Psychological Association, New York City, 1957.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90. (a)

SALZINGER, K., & PISONI, STEPHANIE. Verbal affect responses of schizophrenics as a function of number of reinforcements and rate of acquisition. Paper read at the Eastern Psychological Association, Philadelphia, Pa., 1958. (b)

SALZINGER, SUZANNE. Rate of affect response in schizophrenics as a function of three types of interviewer verbal behavior. Paper read at Eastern Psychological Ass., Atlantic City, March, 1956.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received February 13, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CHANGE IN ATTITUDES TOWARD THE NEGRO IN A SOUTHERN UNIVERSITY

ROBERT K. YOUNG, WILLIAM M. BENSON, AND WAYNE H. HOLTZMAN
University of Texas

A FEW months after the momentous 1954 United States Supreme Court decision concerning segregation an attitude survey was conducted among undergraduate students at the University of Texas (Kelly, Ferson, & Holtzman, 1958). Included in this questionnaire was the 26-item Desegregation Scale (D scale) for measuring attitudes toward the Negro. Background factors found by Kelly et al. (1958) to be significantly related to attitude toward the Negro included the person's geographic region of origin, father's income, major field of study, religious preference, and stated frequency of church attendance.

The existence of these baseline data presented a unique opportunity to study the degree of change, if any, in attitudes toward the Negro after a lapse of several years. The original study had employed a large, stratified random sample of 547 cases which could be compared in detail with an identical sample of undergraduates given the same questionnaire in the winter of 1958. During the intervening three-year period much had happened on the American scene with respect to school desegregation. Stiffening resistance through-out the South came to a dramatic climax in the integration crisis at Little Rock, Arkansas, in late 1957 when Federal troops were called out by President Eisenhower to force the opening of Central High School. The voice of the Southern moderate was drowned out completely by extremists on both sides of the issue.

The period immediately following the Little Rock crisis represented an ideal time for repeating the earlier survey to determine the nature of change in attitudes toward the Negro. The undergraduate student body at the University of Texas is an excellent group for such a study, since cultural mores and life styles characteristic of the deep South, the West, the North, and Mexico are all represented.

PROCEDURE

Great care was taken to duplicate the procedures used in the 1955 survey so that the two sets of data would be comparable. Date cards from the Registrar's Office were used to draw a 5%, random stratified sample of native born, full-time undergraduate students. Stratification was based on sex and religious preference. Of the 629 students to whom letters were mailed, 265 came in immediately to fill out the attitude questionnaire in previously announced group sessions. Another 151 completed the questionnaire after a telephone follow-up. Questionnaires were mailed to the remaining subjects, and 101 of these responded, making a total of 517, or 82% of the original sample (comparing favorably with the 84% obtained in 1955). Preliminary analysis of the data revealed no bias in the questionnaires returned by mail. Since 55 of the 112 nonrespondents had dropped out of school prior to the study, the actual percentage of returns was 91. Eliminating 20 questionnaires that were improperly answered left a total of 497 cases for analysis.

The 90-item attitude questionnaire was given anonymously and scored in the same manner as reported by Kelly et al. (1958). The D scale had been formed from data in a preliminary scaling study by taking the 26 items with minimum variance and approximately equal interval distribution on the 11-point scale. No items were included with scale values ranging from 4.4 to 7.2. Each of the items was scored according to the Likert weighting procedures with extremely positive responses scored zero and extremely negative ones scored four. Summing across all 26 items gave a possible range of scores from zero to 104.

RESULTS

Over-all Comparisons of the 1955 and 1958 Samples

Of paramount interest are differences in the two samples which would indicate any significant shifts in attitude toward the Negro during the three-year interval between 1955 and 1958. The distributions of scores on the D scale were markedly similar in the two samples. The means were 46.2 and 45.3 for the 1955 and 1958 samples, respectively. Similar values for the standard deviations of the distributions were 18.8 and 18.7. Clearly there are no measurable changes in degree of acceptance of the Negro for the undergraduate student body as a whole.

In view of the widespread belief that the so-called moderates have been forced to move away from a neutral position because of the aggressive action of extremists, a comparison of the variances in the two samples is of special interest. The stiffening of extremes should lead to a significant increase in the variance of scores on the D scale. Actually the variance *decreased* slightly from 353 in 1955 to 348 in 1958. These results suggest that the social conflict characteristic of the period from 1955 to 1958 had little or no effect upon attitudes toward the Negro in Texas.

As a further test of changes that might have taken place during the three-year interval between the two surveys, an analysis of variance, corrected for disproportionate cell frequencies, was computed in which the main effects were (*a*) year of survey, (*b*) sex of respondent, and (*c*) fraternity (or sorority) membership. It is quite possible that significant interactions could take place between these factors even though the over-all distributions for the two surveys were markedly similar.

Two significant differences were observed. The first was fraternity affiliation. An F of 4.04 ($P < .05$) was obtained between respondents who belonged to a fraternity or sorority and those who did not. The unaffiliated students tended to be more tolerant of the Negro, a trend that was more pronounced in 1958 than in 1955 (although the sample-by-fraternity interaction failed to prove significant).

The second significant difference was the only one found involving the year of survey as a factor. A significant sex-by-sample interaction ($F = 4.34$, $P < .05$) indicated that the attitudes of men and women changed during the three-year period between surveys but in *opposite directions*. Inspection of Table 1 reveals that men were less tolerant of the Negro in 1958 than in 1955, while women showed a significant increase in tolerance.

TABLE I

CHANGE IN ATTITUDE TOWARD THE NEGRO FROM 1955 TO 1958, MEAN SCORES ON THE D SCALE

	Year of Survey	
	1955	1958
Men	44.8	47.9
Women	45.1	42.3

Additional Analysis of the 1958 Sample

As in the study by Kelly et al. (1958), a complete analysis was made of personal background factors and their relationship to attitude toward the Negro as revealed by the D scale. Since most of these results are in full agreement with the findings obtained for the 1955 sample, they are only briefly summarized. A more extended discussion has already been given in the earlier report. Mean D scale scores are cited parenthetically after the appropriate group.

Geographic location. The area from which the respondent comes is strikingly related to his attitude toward the Negro. In general, tolerance for the Negro increases as one moves from east Texas (mean score of 55.2) to west Texas (45.5) and the Rio Grande Valley (42.9).

Church affiliation. The Baptists (49.8) show the greatest amount of resistance to desegregation, with those stating no religious preference at the opposite extreme (32.4). The Catholics (44.3) and Jews (39.0) are the most tolerant of the major denominations analyzed.

Church attendance. Frequency of church attendance is related to attitude toward the Negro in a curvilinear fashion, again verifying an earlier theoretical deduction by Allport (1954). Those individuals who attend church every Sunday (44.3) and those who never attend church (38.1) are significantly more tolerant than those who claim they attend church once a month (51.4) or twice a month (48.4).

Grade point average. Grade point average is linearly related to D scale scores. Students who are academically superior—B average or above (42.0)—are more acceptant of the Negro than students with grades of C or below (48.6).

Major field of study. Social science majors (35.3) tend to be the most tolerant of the Negro, while students enrolled in pharmacy (57.2) or business (49.9) are the least tolerant.

Size of hometown. Students coming from cities with a population of 50,000 or over (43.3) are more tolerant of the Negro than those who reside in smaller towns and rural areas (48.1). While this same trend was noticed in 1955, it was not sufficiently pronounced at that time to be statistically significant.

Father's income. The higher the stated income, the less tolerant the individual. Students whose fathers had an income of less than $5,000 per year (41.9) were most tolerant, while those who claimed an income of over $20,000 (53.0) were least tolerant.

CONCLUSIONS

Attitude toward the Negro as measured by the D scale appears to be remarkably stable as far as group means are concerned. In view of the social tension prevalent in the South during the period 1955–1958, one might reasonably expect to find a change in attitude toward desegration, at least with respect to a widening of the breach between advocates of different stands on the issue. Contrary to such popular beliefs, no change in either mean score or variance of the D scale was found in comparing a random sample of students at the University of Texas in 1958 with a similar sample in 1955.

The only significant change in attitude over this three-year period is uncovered when men and women in the two samples are analyzed separately. While the women appear to be slightly more tolerant of the Negro in 1958, the men have moved in the opposite direction.

In both the 1955 and 1958 samples, an analysis was made of relationships between personal background characteristics of the student and attitudes toward the Negro. A striking degree of consistency was found across the two samples for such factors as geographic region from where the individual came, church affiliation, stated frequency of church attendance, and major field of study in college. Size of hometown and father's income were also significantly related to attitude score in 1958, although these same trends fell short of significant levels in the 1955 sample.

REFERENCES

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley, 1954.

KELLY, J. G., FERSON, JEAN E., & HOLTZMAN, W. H. The measurement of attitudes toward the Negro in the South. *J. soc. Psychol.*, 1958, **48**, 305–317.

(Received August 17, 1959)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

CRITIQUE AND NOTES

SOME THEORY AND DATA CONCERNING THE ATTENUATION OF AGGRESSION[1]

KENNETH R. WURTZ
Michigan State University

When a child is punished for aggression he gradually learns to anticipate punishment whenever the aggressive drive is activated. This anticipation of punishment, occurring simultaneously with excitation of aggressive behavior, is anxiety. The child learns to be anxious just as he learns most of his other responses, and anxiety comes to be a drive state in and of itself. However, as a drive, anxiety is unique in one respect. When it occurs it is almost always in conflict with other drive states of the organism. In its essence, anxiety is a conflict-producing drive; when it is evoked by aggression, the anxiety-instigated responses conflict with those induced by the child's need to be aggressive.

From a systematic standpoint, anxiety is both an inhibitor and a stimulus. It is an inhibitor of overt forms of aggression by virtue of its capacity to instigate avoidance responses that conflict with aggression, and it is a stimulus to aggression since its occurrence commonly coincides with that of expressed aggression. Since the state of anxiety occurs frequently with the expression of aggression, anxiety itself instigates aggression on the basis of classical conditioning. As a stimulus (apart from its function as an inhibitor) anxiety may be supposed to add instigation to the aggression drive and to facilitate it, since the more relevant stimuli are present the more likely it is that a response will occur. This reasoning leads to the following hypothesis: If the anxiety evoked by a person's aggressive drive is strong enough, it should inhibit the aggressive response; however, if the anxiety is not strong enough to inhibit the aggresive response, the anxiety should facilitate the aggressive response that occurs. In simpler terms, an aggressive response that occurs in the presence of anxiety should tend to be stronger than an aggressive response that occurs in its absence. For example, the more anxious a person is about hitting another person, the harder he should tend to hit him if he does overcome his inhibition against aggression.

The child has to learn to satisfy his aggressive drive, but he has to do it in such a way that it does not call forth aggression anxiety. He must learn to alter his original violent forms of aggression to kinds that do not elicit so much retaliation from his family. One way of doing this is to attenuate the expression of aggression so that the persons against whom it is directed do not perceive the behavior as aggressive. The child learns through experience that there are certain kinds of aggression that the parents do not punish as much as other kinds. Most children are more likely to be punished for aggression that is overtly physical than for nonphysical (e.g., verbal) aggression; they are more likely to be punished for kicking their mother than for wetting the bed or performing some other act that the parents cannot so certainly perceive as aggression. This observation suggests a first principle: the more a child is punished for aggression toward any person, the more attenuated his overt expression of aggression toward that person becomes. The more punishment, the more anxiety, and the more anxiety, the more attenuation.

What happens to the fantasy aggression of the child as a result of the punishment for overt aggression? When aggression is expressed in fantasy, the anxiety associated with aggression against particular objects in real life situations tends to transfer by generalization to the fantasy situation. However, anxiety aroused through such a process of generalization tends to be weaker; hence, in the fantasy situation the aggressive response is more likely to occur than in the reality situation and any accompanying anxiety would tend to facilitate the response. Under these circumstances fantasy aggression is more rewarding for him the more violent it becomes. This reasoning rests on the assumption that direct undisguised aggression most adequately reduces the aggressive drive. A second principle is thus suggested: The more punishment that an aggressive act incurs in real life, the more direct and overt and violent is the aggression that occurs in fantasy.

These two hypotheses derive some support

[1] The writer wishes to express his deepest appreciation to Robert R. Sears, without whose aid this study would have been impossible. The expenses of the study were borne by a grant from the Behavioral Sciences Division of the Ford Foundation.

from the results of a study by Hollenberg and Sperry (1951). In the first part of their study, the experimenter verbally punished one group of children and did not verbally punish another group for showing aggression during doll play. The condition of verbal punishment is more similar in this respect to everyday life than to fantasy, whereas the condition of no verbal punishment represents a "pure" fantasy situation. Under the condition of verbal punishment, consequently, anxiety should act as an inhibitor and the expressed aggression be more attenuated (less "intense" in Sperry's terms) in the verbally punished group than in the control group; the data tend to support this prediction. Since Hollenberg and Sperry did not present the raw data in their paper, it is impossible to determine if this difference is significant.

In the second part of their research, a different set of *Ss* was divided into four groups according to experiences at home: those who had (*a*) low punishment for aggression and low frustration, (*b*) low punishment and high frustration, (*c*) high punishment and low frustration, and (*d*) high punishment and high frustration. These groups were compared with respect to the percentage of fantasy aggression that was "intense." In this study, the experimenter did not punish aggressive responses; consequently, the doll play responses represent fantasy responses and anxiety does not tend to be strong enough to act as an inhibitor. To test this hypothesis with frustration held constant, Group *a* is compared with *c* and Group *b* with Group *d*. In both comparisons, the hypothesis is that the group with the high punishment for aggression (and therefore high aggression anxiety) should show less attenuated (more "intense") aggression than the group with the low punishment; the data support this prediction, but again it is impossible to determine the significance of the differences.

If one assumes that more anxiety is aroused when aggression is directed toward adults than when aggression is directed toward other children, then it may be further hypothesized that in doll play in which aggression is not punished, the child's aggression toward the parent dolls tends to be more violent and direct (since it has the added instigation of anxiety) and that his aggression toward child dolls tends to be more attenuated. If aggression is considered to be more violent and less attenuated when it involves "bodily injury" than when it does not, one should expect a smaller proportion of aggressive responses involving bodily injury with child doll objects than with adult doll objects.

Although the same prediction is made about the doll agents of aggression, the reasoning behind the prediction is somewhat different than that underlying the prediction about doll objects. The objects of aggression in doll play represent the persons toward whom the child would like to direct aggression in real life; the agents of aggression in doll play presumably refer to the persons that the child would like to name as being aggressive. The child's parents punish him and this is aggression. However, the parents also punish him for saying that they are aggressive. Consequently, in real life situations anxiety is aroused whenever the child attributes aggressive acts to the parents. The expectation is therefore that a greater proportion of aggressive responses in doll play should be of the nonbodily injury type with child doll agents than with adult doll agents.

SOME RELEVANT DATA

The data to test this hypothesis were taken from a study previously reported by Pauline Sears (1951) and by Sears, Pintler, and Sears (1946). In this study, 150 children, 75 boys and 75 girls, 25 of each sex in the age groups three, four, and five years were given two 20-min. sessions of doll play with a standard doll family. The children were allowed complete freedom of response, and the experimenter did not punish aggressive behavior. Consequently, the child's responses in this study may be viewed as fantasy responses. Acts of aggression were recorded in detail and later scored in one of the following five categories: Type 1—the doll was used as an agent of bodily injury aggression; Type 2—the doll was used as an object of bodily injury aggression; Type 3—the doll was used as an agent of nonbodily injury aggression; Type 4—the doll was used as an object of nonbodily injury aggression; Type 5—all other forms of aggression.

To test the above hypothesis with the agents of aggression, an index of attenuation was calculated for each of the five dolls in each session by the formula: Type 3 ÷ (Type 1 + Type 3). (Children who did not show aggression of either type were omitted in computing the index.) Substituting Type 2 for Type 1 and Type 4 for Type 3 the same formula was used for objects of aggression. The higher these percentage values, the more relatively attenuated was the total aggression.

The significance of these differences was tested by calculating the CRs of the differences between the summed percentages of attenuated aggression for the two parent dolls with those for the three child dolls. These comparisons are shown in Table

TABLE 1
RELIABILITY OF DIFFERENCES BETWEEN SUMMED PERCENTAGE OF ATTENUATED AGGRESSION

	Agents						Objects					
	Session I			Session II			Session I			Session II		
	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls	Parent Dolls		Child Dolls
Boys	.54		.63	.43		.65	.28		.30	.20		.23
CR		.88			2.16			.22			.40	
P		NS			.03			NS			NS	
Girls	.69		.83	.54		.78	.37		.52	.24		.44
CR		1.10			2.31			1.34			1.97	
P		NS			.02			NS			.05	
Total[a]	CR = 3.10 *P* = .001						CR = 6.67 *P* = .0001					

[a] Represents the sum of males and females for both sessions.

1; two-tailed tests were used in computing the *P* values. The totals for agents and objects refer to the sums over both sexes and sessions. The order of events was as follows: the agents were tested for attenuation, the present theoretical explanation of the results was developed, and the prediction based on it then tested on the objects.

These results support the hypothesis of this paper fairly well, considering the crudeness of the measure of attenuation used. In fantasy as compared with real life, anxiety functions more as a stimulus than as an inhibitor.

REFERENCES

HOLLENBERG, E., & SPERRY, J. Some antecedents of aggression and effects of frustration in doll play. *Personality*, 1951, **1**, 32–43.

SEARS, PAULINE S. Doll play aggression in normal young children: Influence of sex, age, sibling status, father's absence. *Psychol. Monogr.*, 1951, **65**(6, Whole No. 323).

SEARS, R., PINTLER, M., & SEARS, PAULINE S. Effect of father separation on preschool children's doll play aggression. *Child Developm.*, 1946, **17**, 219–243.

(Received October 3, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

INDIVIDUAL DIFFERENCES IN REACTION TO FAILURE-INDUCED STRESS[1]

SAMUEL C. FULKERSON[2]
USAF School of Aviation Medicine

IN THE PAST a number of attempts have been made to find personality correlates with behavior under stress. Lazarus, Deese, and Osler (1952), reviewing the work up to 1951, indicated that scarcely any results of practical import had been found and emphasized the possibility of individual differences as a primary determinant of the stress reaction.

Since this review, several studies have appeared relating performance under stress to Janet's hysteria-psychasthenia personality typology. For instance, Eriksen (1954b) found that *Ss* with higher scores on the Hysteria (*Hy*) scale of the Minnesota Multiphasic Personality Inventory showed greater stimulus generalization under stress than *Ss* with higher scores on the Psychasthenia (*Pt*) scale. Deese, Lazarus, and Keenan (1953) reported results which suggest that learning under stress may be related to these same MMPI scores.

[1] This study is part of a project on the adaptability screening of flying personnel. Appreciation is expressed to Saul B. Sells and David K. Trites for their suggestions and criticisms, and to Wayne H. Holtzman for making available the data from the ROTC sample.

[2] Now at Western Psychiatric Institute, University of Pittsburgh Medical School.

The present study, conceived within the same frame of reference, is an attempt to relate scores on the *Hy* and *Pt* scales of the MMPI to performance under stress where the experimental task involves a kind of behavior which has been said to be important in distinguishing the hysteric personality type from the psychasthenic. The task was the McKinney Reporting Test (McKinney, Strother, Hines, & Allee, 1951), which introduces at its midpoint a psychological stress condition designed to elicit the threat of failure. The test requires *S* to count a number of small details in an accurate manner.

Insofar as the stress conditions arouse habitual defense reactions, the tendency for the hysteric to be careless about details and for the psychasthenic to be meticulous about them should be accentuated. The first hypothesis, then, was that psychasthenics would become more accurate than usual under stress and hysterics less accurate.

It also seemed likely that defense reactions would be stronger and more consistent for a poorly adjusted person than for a well adjusted person. Therefore, the second hypothesis was that the postulated differences in accuracy of performance between hysterics and psychasthenics would be greater the more poorly adjusted the individual.

METHOD

Subjects

For several years the USAF School of Aviation Medicine has supported a large scale research program on pilot selection. From this program four samples were available which had been administered the McKinney Reporting Test and the MMPI. Sample A consisted of 475 aviation cadets who were tested during the primary phase of pilot training. Sample B consisted of 428 copilots, tested during a three-month course in Combat Crew Training. Sample C consisted of 325 aircraft commanders, also tested during the course in Combat Crew Training. Sample D contained 118 ROTC officer cadets at the University of Texas.

Adjustment ratings of different kinds had been made within each of these samples. In Samples A, B, and C, adjustment was estimated primarily on the basis of a psychological examination, including an interview by a clinical psychologist, and the results from several projective tests. (Neither the MMPI nor the McKinney test were available to the raters.) In addition, training records, including medical records, demerits, academic grades, buddy ratings, and instructor ratings were also used to decide the adjustment of the trainee.

On the basis of these measures, two extreme groups were selected within each sample, one consisting of cases who showed clear-cut evidence of inability to adapt to the training program and who were also given a low adjustment rating by the psychologist. The other group contained individuals who had outstanding records during training as well as being rated high on adjustment.

From Sample A, 100 low adjustment and 100 high adjustment cases were selected to form Group A. Of Sample B, 93 low and 97 high adjustment cases formed Group B. From Sample C, 94 low and 97 high adjustment cases were selected for Group C. In this study the analysis was based on these groups rather than the total samples.

Sample D was treated differently because of its small size. Adjustment was measured by buddy ratings alone. Each *S* was asked to rank order all the men in his ROTC unit whom he knew well on a number of variables, including "personal adjustment." An average "personal adjustment" rating was computed on each individual and used as the adjustment measure. These ratings were then divided at the median, with the 58 cases below the median designated as low adjustment *S*s and the 60 above it as high adjustment *S*s.

Measures

The McKinney test consists of three parts: a prestress period, in which *S* can work at his own speed; a stress period, in which he is pressured to work beyond his capacity under the threat of failure; and a final poststress period, which returns to the conditions of the prestress period. Only measures from the prestress and stress periods were used in this study. The items on the test consist of diagrams containing small circles and lines, and the task is to count the number of these circles and dashes in each diagram. For the prestress and stress periods, the number of items attempted and the number of items correct were calculated. For each period, an accuracy score was computed by dividing the number of items correct by the number of items attempted. This gave the percentage of items correct. These percentage scores were transformed to standard scores.

In order to test the first hypothesis, that the threat of failure arouses the defense mechanisms typical for the hysteric or psychasthenic, a difference score was obtained between the accuracy scores by subtracting the score for the stress period from that for the prestress period. A constant was added to get rid of negative numbers. The *Hy* and *Pt* scales of the MMPI were then correlated with the McKinney difference scores.

Earlier studies (Fulkerson, 1957; Fulkerson, Freud, & Raynor, 1958) revealed that among pilots and pilot trainees, the *K* score was posi-

tively correlated with success in training, and its use as a "suppressor variable" decreased the discriminability of the scales. Therefore, the raw *Pt* scores, uncorrected by *K*, were converted into *T* scores. The *T* values used to convert *Pt* and *Hy* to standard scores were based on pilot norms, because the original norms were known (Fulkerson et al., 1958) to be inappropriate for a pilot sample.

RESULTS

The Pearson product-moment correlation coefficients between the McKinney and MMPI scores are shown in Table 1. An average of the coefficients for the four groups was obtained by the method outlined by Edwards (1950, p. 133). The correlations involving *Pt* were not significantly different from zero. The *Hy* score correlated significantly with the McKinney score in the predicted direction for the pooled low adjustment *Ss*, with those having high *Hy* scores tending to shift toward greater inaccuracy, relative to the rest of the group, on being exposed to the stress period on the McKinney test.

The second hypothesis predicted stronger relationships for the low adjustment *Ss*. Again, the correlations involving *Hy* are in line with this hypothesis.

The difference score derived from the McKinney test was designed to partial out the effect of initial differences in accuracy level from the accuracy score obtained on the stress period. It can be asked whether this gross correction did in fact improve the strength of the hypothesized relationships. Table 2 presents the correlations, averaged over the four groups, for the prestress and stress period accuracy scores on the McKinney test. The correlations for the prestress period are uniformly insignificant, and those involving the stress period are similar to those based on the difference score, though not consistently in the predicted direction.

TABLE 1

CORRELATIONS BETWEEN SHIFT SCORE ON THE MCKINNEY REPORTING TEST AND THE MMPI SCALES OF *HY* AND *PT*

Group	Low Adjustment Subjects			High Adjustment Subjects		
	Hy	*Pt*	*N*	*Hy*	*Pt*	*N*
A	.17	−.03	100	−.22*	.10	100
B	.16	.04	93	−.05	−.19	97
C	.09	−.10	94	.09	−.13	97
D	.26*	.06	58	−.27*	.11	60
Average *r*	.16**	−.02		−.10	−.04	

* Significant at .01–.05 level.
** Significant at less than .01 level.

TABLE 2

AVERAGE CORRELATIONS BETWEEN THE ACCURACY SCORES ON THE PRESTRESS AND STRESS PERIODS OF THE MCKINNEY REPORTING TEST AND THE MMPI SCALES OF *HY* AND *PT*

McKinney Test	Low Adjustment Subjects		High Adjustment Subjects	
	Hy	*Pt*	*Hy*	*Pt*
Prestress period	.03	−.04	.07	−.07
Stress period	−.13*	−.03	.17**	−.02

* Significant at .01–.05 level.
** Significant at less than .01 level.

DISCUSSION

The results of this study can hardly be said to demonstrate striking individual differences. They suggest that unless quite large samples are used, which contain a high proportion of poorly adjusted *Ss*, the hypothesized relationship between performance under stress and the hysteria-psychasthenia dimension is not likely to appear. Even taking these factors into account, the personality dimension, at least as measured by the *Hy* scale, identified only 3% of the variance involved in the performance under stress. Of course *Hy* is not necessarily an accurate measure of the personality dimension, nor the McKinney test necessarily one that educes a powerful stress reaction. Another point to keep in mind is the homogeneity of these samples on several variables. All *Ss* were relatively young, the ROTC student officers and the aviation cadets being primarily in their early twenties, the aircraft commanders and copilots in their late twenties and early thirties; all were above average in intelligence; there was homogeneity of vocational interests among the military samples; and, at least for the military samples, there was a strong set to do well on these tests, in all probability, since these *Ss* had no reason to suppose the results might not affect their careers in some way. In a more heterogeneous sample, individual differences in reaction to stress might be clearer.

The results also indicate that *Hy* and *Pt* are not measures of a single bipolar trait. Eriksen (1954a) has reported similar findings and concluded that *Hy* is measuring the hysteric-psychasthenia dimension, whereas *Pt* seems to measure neuroticism or ego strength. In the present study, the over-all correlation between *Hy* and *Pt* was −.09. This correlation, however, was a function of adjustment level. For the low adjustment cases, the correlation was .06; for the high adjustment cases, it was

−.24. This latter value is significant beyond the .01 level.

These differences, as well as those shown in Table 1, indicate that adjustment level is a variable which must be controlled along with such standard variables as age and intelligence in personality studies.

SUMMARY

The McKinney Reporting Test and the MMPI were administered to samples of aviation cadets, aircraft commanders, copilots, and ROTC student officers. Two criterion groups representing the extremes in adjustment were selected within each of the military samples to form high and low adjustment groups. The ROTC officers were divided at the median into upper and lower adjustment groups.

There was an over-all significant tendency among the low adjustment cases for those who shifted toward more accurate performance on the McKinney stress period to answer on the *Hy* scale of the MMPI like Janet's psychasthenic type, and for those who became more inaccurate to answer like the hysteric. This tendency did not hold for high adjustment cases. The McKinney test did not correlate with the *Pt* scale.

REFERENCES

DEESE, H., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

ERIKSEN, C. W. Psychological defenses and "ego strength" in the recall of completed and incompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45–50. (a)

ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561–564. (b)

FULKERSON, S. C. Adaptability screening of flying personnel: Research on the Minnesota Multiphasic Personality Inventory. *USAF Sch. Aviat. Med., Rep.*, 1957, Proj. No. 57–106.

FULKERSON, S. C., FREUD, S. L., & RAYNOR, G. H. The use of the MMPI in the psychological evaluation of pilots. *J. aviat. Med.*, 1958, **29**, 122–129.

LAZARUS, R. S., DEESE, J., & OSLER, S. F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MCKINNEY, F., STROTHER, G. B., HINES, R. R., & ALLEE, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316–323.

(Received October 27, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

A CROSS-CULTURAL STUDY OF STRUCTURED GROUPS WITH UNPOPULAR CENTRAL MEMBERS

A. I. MOHANNA
Al-Azhar University, Cairo

AND MICHAEL ARGYLE
University of Oxford

THIS paper reports an experiment on the performance of small social groups with restricted channels of communication. The method used by Leavitt (1951) was followed, and the following additional variables were manipulated: (*a*) the central member of the wheel groups was selected sociometrically to be popular or unpopular (the validity of this manipulation being confirmed by a postexperimental questionnaire), (*b*) the experiment was repeated in Oxford and in Cairo, with all other conditions held as constant as possible. Altogether 80 groups were used, involving 400 schoolboys aged 13½–15½.

Results

Wheel vs. circle patterns. Leavitt's result was confirmed that the wheel pattern is superior in time taken, number of messages required, and errors made (see Table 1).

Wheel pattern with popular vs. unpopular central members. As can be seen in Figures 1 and 2, the unpopular wheel groups settle down to a level of performance better than the circle groups, and worse than the popular wheel groups, in time taken and messages used, and "learn" faster than the other two kinds of groups.

Oxford vs. Cairo. Table 1 indicates that there

TABLE 1
WHEEL V. CIRCLE PATTERN—RESULTS FOR TRIALS 17–24

		Oxford		Cairo	
Time taken (secs.)	Wheel	43.1		32.5	
	Circle	103.1	($p < .01$)	83.4	($p < .001$)
Number of messages	Wheel	10.0		8.1	
	Circle	13.9	(n/s)	16.0	($p < .001$)
Number of errors	Wheel	0.78		0.34	
	Circle	1.31	(n/s)	0.92	($p < .01$)

Note.—$n = 8$ for each condition. The t test was used throughout, with a one-tailed distribution.

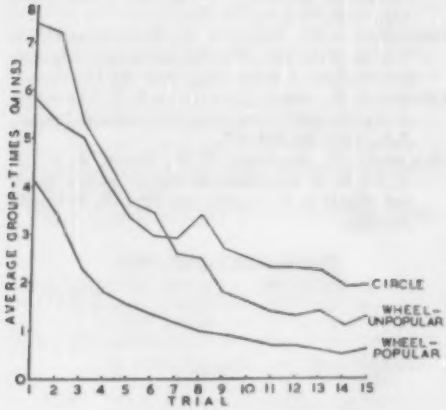


FIG. 1. AVERAGE GROUP TIMES PER TRIAL (16 GROUPS ON EACH CONDITION)

are only slight (and insignificant) differences in absolute efficiency between the Oxford and the Cairo groups. Satisfaction was greater in popular wheel groups, followed by circle groups and unpopular wheel groups; this result was found in Cairo only, but the generally higher level of satisfaction of the Oxford *Ss* may have concealed

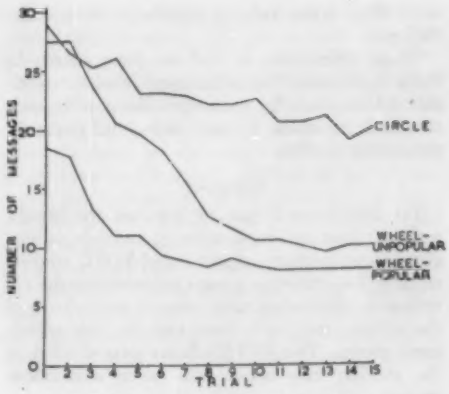


FIG. 2. AVERAGE MESSAGES PER TRIAL (16 GROUPS IN EACH CONDITION)

any differences. Otherwise the results in Oxford and Cairo were closely similar.

Discussion

The inefficiency of unpopular compared with popular wheel groups may be due to the same causes as the low output of noncohesive groups on cooperative group tasks—that there is an additional incentive for work (Argyle, 1957, pp. 126–8). The more rapid learning in unpopular wheel groups could be due to a short-term version of "interaction leads to liking"; that is, the groups are extremely slow at first, but greater speed becomes possible as the initial hostility is *partly* reduced by interaction.

REFERENCES

ARGYLE, M. *The scientific study of social behaviour.* London: Methuen, 1957.

LEAVITT, H. J. Some effects of certain communication patterns on group performance. *J. abnorm. soc. Psychol.*, 1951, **46**, 38–50.

(Received September 4, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

RECALLED DREAM THEMES AND PERSONALITY[1]

JOSEPH F. RYCHLAK
Washington State University

ALTHOUGH ACCEPTED as legitimate data for personality analysis by most clinicians, dreams have only occasionally been studied systematically in relation to personality variables. The present study of children's dreams was designed to test the hypothesis that persons dream in a pat-

[1] The author would like to thank the following people for their help with the study: J. G. Wynn Jr., Mary C. Witt, and J. H. Diamond of University School, Florida State University; C. Moos of the Psychology Department, Florida State University; Alice O. Thomas of Pullman, Washington.

terned fashion and that such uniformities can be related to objective test findings, and possibly to sociometric ratings. The fact that these dreams are "recalled" is given at the outset. Paralleling the clinical procedure of weekly sessions with a client, the study concerned dreams recalled by *S*s over the past week.

METHOD

Dream Measurement

Each dream collected was given a single scoring, according to its major theme. A major theme was seen as reflecting the manifest point, or story line of the dream, which usually was easily identified because *S*s reported only the core of their dreams. In longer and more complex reports the theme scored as major usually revolved about the final event of the dream story.

Dreams were scored for three major themes: Affiliative, Reward, and Tension. A method of scoring was devised on the basis of previous research findings (Alexander & Wilson, 1935; Blanchard, 1926; Jersild, Markey & Jersild, 1933; Witty & Kopel, 1939) and preliminary exploration. A brief description of each category with examples taken from *S*s' reports follows:

Affiliative: Dreams having relaxed or pleasurable interpersonal relations with other human beings as their major themes, regardless of the latter's relationship to *S*; e.g., relatives, peers, same or opposite sex, etc. There must be some form of interaction with the person, and not simply a recollection of seeing him in the dream.

Examples:

I dreamed that my brother and I went swimming.
I dreamed that a boy in our room took every girl out to a show.
I dreamed that me and Stevie played catch.

Reward: Dreams with major themes involving a positive, pleasurable connotation *other than* that derived from strictly interpersonal contact. Reward themes include individual achievement and recognition, the receiving of gifts, and any phantasy-like tale of miraculous happenings.

Examples:

I dreamed I got a new watch.
I dreamed I scored 25 points against Panama City.
I was on Arthur Godfrey's Talent Scouts.

Tension: Dreams reflecting anxiety, frustration, or hostility in their major themes.

Examples:

I dreamed that Frankenstein was after me.
I was riding my bike and the wheel came off.
I dreamed some friends of mine were killed in an auto accident.

Less than 5% of the dreams were unscorable. A reliability check was performed on 32 randomly chosen dreams. The *E* and a trained rater independently scored each dream with 94% agreement (agreements divided by total comparisons).

Personality Measurement

Cattell's Junior Personality Quiz (Cattell, Beloff, Flint, & Gruen, 1953) was used, a 12-factor test applicable to the age range of 10 to 16 years. It contains 144 items and takes approximately one hour to administer.

Popularity Measurement

Popularity was assessed sociometrically. Each *S* selected three children (either sex) with whom he would "like best to play any of the games that we know" and three children with whom he would "like to work or make something." The popularity score was the absolute number of choices an *S* received from his classmates.

Subjects

Two elementary school classes were studied, a fifth grade (17 females, 13 males) and an eighth grade (15 females, 14 males). The final sample of children who reported at least two dreams included 25 girls (13 fifth, 12 eighth graders) and 17 boys (7 fifth, 10 eighth graders). In general the children had middle-class backgrounds. Mean CAs were 10-9 for the fifth graders and 14-1 for the eighth graders. Mean grade-level scores for Total Language Ability on The California Achievement Test (Tiegs & Clark, 1951) were as follows: fifth graders 5.99, eighth graders 8.87.

Procedure

Dream themes were collected in writing on the same day of each week over a three-month period. When more than one theme was recalled it was accepted. If a child could not recall having a dream over the past week he was asked to indicate this in writing. If he was absent on the day of collection he reported his dream on the first day of his return. Emphasis was repeatedly placed on honesty in reporting, and reassurance was given that not all children have dreams and many who do sometimes forget them. After 12 weeks each *S* was requested to make up a "daydream" in class.

Data obtained in this way have obvious limitations. A child might make up a dream story to please authority figures. *S*s with poor verbal skills may neglect reporting dreams because of the effort it takes to record them. To check on these factors certain measures were taken.

First, the daydreams were used as a form of

TABLE 1
DREAM PRODUCTIVITY AND MAJOR THEMES
(N = 59)

Sample Breakdown	Dream Incidence		Percentage of n reporting one or more themes of		
	Median	Med. Devia.	Affiliative	Reward	Tension
All Boys (n = 27)	2.20	1.48	26*	59	67
All Girls (n = 32)	2.90	2.33	56	44	72
5th Graders (n = 30)	2.33	2.22	40	57	67
8th Graders (n = 29)	2.62	1.83	45	45	73

* Significantly different than comparable percentage for girls: P = .02.

check on the presumed legitimate dream data. It was found that these imagined dreams were consistently longer, generally more coherent and logical in their themes, and much more likely to have well defined, happy endings. On the basis of such principles a trained rater checked "actual" or "daydream" following the reading of 10 dreams of each type randomly selected, typed on plain paper and shuffled before rating. He was able to discriminate correctly in 75% of the cases (P = .042).

Secondly, comparisons were made between the number of dreams submitted by *Ss* and their scores on two indices presumably related to verbal skills: intelligence as measured by the personality test (Factor 12, Cattell et al., 1953) and Total Language Scores of The California Achievement Test (Tiegs & Clark, 1951). These chi squares failed to reach even the .10 level of significance.

RESULTS

Table 1 reports the median number of dreams collected, and the percentage of *Ss* producing at least one theme in each of the three dream categories. Thus, 26% of the boys and 56% of the girls reported at least one Affiliative theme over the 12 weeks of the study, the only significant difference in the table. Considering only those *Ss* who recalled two or more dreams (to whom subsequent analysis was restricted by prior decision) the median values became: girls (n = 25), Median 3.44, Median Deviation 2.30; boys (n = 17), Median 3.20, Median Deviation 1.15. In order to assess possible influences of pubescence on major themes, the Affiliative dreams were rescored on the basis of obviously romantic content (e.g., "I dreamed that Johnnie kissed me"); 25% of the fifth graders and 27% of the eighth graders reported from one to four heterosexual themes.

The *Ss* who produced two or more dreams were ranked with respect to the proportion of reported dreams classified under each theme. In the ensuing analysis, the sample was dichotomized at the median on each measure under consideration, and null hypotheses of independence evaluated by chi square.

Dreamers falling above the median on Affiliative themes fell below the median on sociometric play choices (chi square, 4.71, P = .04). Reward dreamers tended, but not significantly so, to fall above the median in play choices (chi square, 2.24, P = .15). On the work group choices the Affiliative dreamers once again fell below the median in popularity (4.68, P = .04) and the Reward dreamers showed a comparable tendency to popularity (2.40, P = .13). Tension dreamers failed to show significant consistencies of choice in either group (chi squares of 1.61 and .088). When the choices for both work and play groups are considered as a single sociometric score, the chi squares increase as follows: Affiliative themes and low popularity 7.78 (p < .01); Reward themes and tendency to popularity 2.66 (P = .11).

On the personality variables, seven comparisons were accepted as reflecting a significant difference. Affiliative themes bore a relationship to the following three personality variables: Emotional Sensitivity (9.54, P < .01), Schizothymia (4.71, P = .04), and Socialized Morale (4.71, P = .04). Reward dreamers fell within the personality variables of Cyclothymia (3.38, P = .07) and Independent Dominance (12.30, P < .001). Tension dreamers scored high on the variables of Withdrawn Schizothymia (4.87, P = .04) and Lowered Intelligence (4.71, P = .04).

DISCUSSION

The hypothesis under consideration has received some verification. Manifest dream themes have been shown to discriminate fairly well between *Ss* and to relate to personality and sociometric measurements. Two of the factors bearing a significant relationship to dreams are among the least reliable of the Personality Quiz (split-half r): Cyclothymia vs. Schizothymia (.31) and Independent Dominance (.30). Interpretations made on the basis of these variables must be considered tentative. On the other hand, the two most reliable factors of the test also entered into the findings: Emotional Sensitivity (.83) and Intelligence (.78).

In spite of the gross classification employed in this study, a review of the findings in conjunction with the definitions of the measured personality factors (Cattell et al., 1953, pp. 8–11) suggests certain common personality trends among dreamer

types. Passive-conformity, inhibition, and a preference for close interpersonal relations seem to typify the Affiliative dreamer. This motivational pattern may be viewed as either cause or result of the relatively low popularity disclosed by the sociometric data. The Reward dreamer seems to be either outgoing and affable or fairly dominant and achievement-oriented. Needs reflected in Reward themes are not inconsistent with these behaviors. *Ss* having many Tension dreams scored highest on one of the factors reflecting neurotic trends, as might be expected from clinical experience. Their lower standing on intelligence may reflect a relationship between intelligence and frustration in school tasks.

REFERENCES

ALEXANDER, F., & WILSON, G. W. Quantitative dream studies. A methodological attempt at quantitative evaluation of psychoanalytical material. *Psychoanal. Quart.*, 1935, **4**, 371–407.

BLANCHARD, P. A study of subject matter and motivation of children's dreams. *J. abnorm. soc. Psychol.*, 1926, **21**, 24–37.

CATTELL, R. B., BELOFF, J., FLINT, D., & GRUEN, W. *Handbook for the junior personality quiz.* Champaign, Ill.: Inst. for Person. & Ability Testing, 1953.

JERSILD, A. T., MARKEY, F. V., & JERSILD, C. L. Children's fears, dreams, wishes, daydreams, likes, dislikes, pleasant and unpleasant memories. *Child Development Monographs*, 1933, No. 12.

TIEGS, E. W., & CLARK, W. W. *Manual of directions for the primary California achievement test battery.* Los Angeles: California Test Bur., 1951.

WITTY, P. A., & KOPEL, D. The dreams and wishes of elementary school children. *J. educ. Psychol.*, 1939, **30**, 199–205.

(Received November 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

SCHIZOPHRENIC-LIKE RESPONSES IN NORMAL SUBJECTS UNDER TIME PRESSURE[1]

GEORGE USDANSKY AND LOREN J. CHAPMAN[2]

University of Chicago

A NUMBER OF THEORISTS and investigators have discussed the formal similarities of the thinking of schizophrenics to that of normal *Ss* under special conditions. Soon after Freud described primary process thinking in dreams, Jung pointed out similarities between dream thinking and the thinking of schizophrenics. Jung also reported similarities on his word association test between schizophrenics and normals in states of drowsiness or reduced attention. Other conditions that have since been reported as producing schizophrenic-like thinking in normals include organic brain damage, drug induced states, oxygen deprivation, personalized distraction, fatigue, and time pressure. Cameron and Magaret (1951) add several more to the list, including interruption of ongoing activity, environmental changes, preoccupation, and emotional excitement.

It appears easy to fit atypical normal thinking into categories that have been used to describe schizophrenic thought, perhaps in part because of the vagueness of the categories. In drawing such comparisons—or in drawing any conclusions about thinking pathology—it is important to specify exactly the nature of the thinking deviation. The present investigation of a condition under which normal *Ss* show an increased tendency to make a schizophrenic kind of error uses a method that makes possible precise specification of the nature of the error. In place of using the traditional method of observation followed by after-the-fact classification, *E* administered a concept sorting task that has built into it opportunities for making specifically associative errors. The task has previously been shown to distinguish a schizophrenic and a normal group by the number of incorrect responses that are associative in nature (Chapman, 1958). Such errors appear on the basis of clinical evidence to be important in the thinking deviation of schizophrenics (Arieti, 1955; Bleuler, 1950; Jung, 1936; Kraepelin, 1919).

The special condition investigated here was that of speeding up responses by instructing *S* to hurry. It was hypothesized that speeding up the responses of normals would produce an increase in associative errors.

METHOD

Procedure. The *S* was asked to place a sorting card bearing the name of a common object, e.g.,

[1] This study was supported in part by a research grant (M-1416) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Now at the University of Kentucky.

TABLE 1
EXAMPLES OF TEST ITEMS

Sorting Card Word	Guide Card Words[a]		
gold	typewriter (I)	steel (C)	fish (A)
hat	head (A)	dress (C)	garage (I)
raincoat	pencil (I)	water (A)	awning (C)
lip	stick (A)	ear (C)	chair (I)
dragon	rope (I)	trout (C)	wagon (A)
face	lace (A)	weed (I)	knee (C)

[a] (C) identifies a word as the correct conceptual choice, (A) the associative distracter, and (I) the irrelevant. See text for further explanation.

"gold," with that one of three guide card words which "could be called by the same name or term" as the sorting card word. In the "gold" example, the three guide card words are: typewriter, steel, and fish. The correct answer is "steel" because both it and "gold" can be called "metal." "Fish" is an associative distracter on the basis of commonly experienced contiguity of words. "Typewriter," on the other hand, is a word intended to have neither conceptual closeness nor associative connection with the sorting word. It is called here an "irrelevant" word. Two other types of associative distracters were used, associations based on contiguity of objects in common experience, e.g., hat–head, raincoat–water, and those based on rhymes, e.g., dragon–wagon, face–lace. Table 1 lists six of the items. Increased speed of sorting could be expected to result in an increase in errors of all types. But errors other than in response to the associative distracter words should, by chance, be equally divided among placements with the associative guide card word and the irrelevant one. These errors contribute, in proportion to their number, to raising the total of associative errors. Therefore, a measure of the effect of associative pull alone is a score of associative placements minus the irrelevant placements. This association-minus-irrelevant score should yield a comparatively pure measure of response to the associative connections. Since an earlier study showed that this error score is responsible for much of the inferiority of schizophrenics to normals on this task, an increase of this score in normals under time pressure may be described as an increase in schizophrenic-like errors.

Ninety test items like those described were used. The 90 were divided into two blocks of 45, with each of the three bases of association equally represented in each block. Each *S* sorted one block of cards under the fast condition and the other under the slow. The two blocks appeared equally often in the two conditions. To control for the differential effects of practice and fatigue, each of the two experimental conditions was equally often presented first and second.

Instructions. The *S* was instructed: "For each card in your hand, you decide which word up here mentions something that can be called by the same name or term, and you place your card in the correct tray." This instruction was followed by six examples accompanied by further explanation. The *S* who was to sort under the slow condition first was then told: "Now go ahead on your own with the rest of the cards in the pile before you. Your speed is not important. I want you to take your time and do your best on each card." After completing the first block of cards, he was given the alternate block with instructions stressing that he was to sort these at his topmost speed, that the interest on this block of cards was in "just how fast you can get through these cards while still trying to get them right." Instructions were substantially the same, but in reverse order, for those *S*s who sorted under the fast conditions first.

Apparatus. Sorting cards were 3 × 5. Each guide card, 2 × 11, had its three guide card words printed on a single line. The trays were of cardboard, 5⅛ deep and 7 square. The guide cards were placed on the table so that they were directly in *S*'s line of vision, with a tray in front of each guide card word. *E*, holding the guide cards on the table, sat opposite *S* and removed the top card just before *S* completed his placement in one of the trays. Thus there was continually a new guide card confronting *S*.

Subjects. The *S*s were 28 junior college and teacher's college male students ranging from freshmen to seniors and in age from 18 to 41 years, with a mean age of 24.4 years. Three additional *S*s were tested but were not included in the analysis of results because the differences between their "fast" and "slow" sorting times were atypically slight, indicating that the *S*s were not following the instructions. These atypical differences (20, 17, and 4 sec.) were all more than 2 *SD* units from the group mean difference.

RESULTS

The *S*s responded to the instructions for fast and slow sorting, as shown by a mean sorting time of 114 sec. under the fast condition and 220 sec. under the slow.

Table 2 lists the means and medians for the various error scores under both fast and slow conditions, showing scores separately for the three bases of association as well as for combined association and combined irrelevant scores. The asso-

TABLE 2
ERROR SCORES UNDER THE FAST AND SLOW CONDITIONS

Score and Associative Distracter	Fast Condition		Slow Condition		Difference (Fast minus Slow)		
	M	*Mdn*	*M*	*Mdn*	*M*	*Mdn*	*CR*
Association choices							
Rhyme	1.93	1.10	.68	.41			
Word	3.21	2.30	1.93	1.20			
Object	4.89	4.50	3.43	1.25			
Combined	10.04	9.00	6.04	2.50	4.00	3.50	3.12
Irrelevant choices							
Rhyme	.75	.50	.14	.28			
Word	.71	.60	.07	.27			
Object	.39	.35	.04	.26			
Combined	1.86	1.39	.25	.30	1.61	1.36	3.90
Association minus Irrelevant							
Rhyme	1.18	.30	.54	.25	.64	.13	.98
Word	2.50	1.83	1.86	1.17	.64	.83	1.74
Object	4.50	4.00	3.39	1.25	1.11	1.00	1.49
Combined	8.18	6.83	5.79	2.50	2.39	2.00	2.25

ciative basis of any given irrelevant error is determined by the type of association error possible on that same item.

Because of the strongly skewed distribution of both associative and irrelevant error scores, a Wilcoxon signed rank test was used to test the significance of the difference between various kinds of errors. All values were computed using a two-tailed test. The associative error scores exceeded the irrelevant error scores ($p < .01$) for both the fast and slow conditions.

The third section of Table 2, headed Fast minus Slow, contains the means, medians, and critical ratios for the differences between conditions. Taking the three bases of association together, we see that, as expected, both the associative and the irrelevant error scores increased under the fast condition ($p < .01$). Of greatest interest, however, is the increase under the fast condition of the association-minus-irrelevant score ($p < .03$). This score, it will be recalled, is the purer measure of associative intrusion. We conclude that normal performance under time pressure was significantly impaired due to an increase in associative errors.

Table 2 further shows that all three bases of associative distraction contributed to this significant difference, between conditions, of the total association-minus-irrelevant score. Taken separately, however, none of the three attained significance.

DISCUSSION

This experiment has shown that certain errors made by normals under time pressure are like the errors of schizophrenics. Chapman (1958) has recently given evidence that the occurrence of schizophrenic-like errors in normals is not dependent on "special conditions," but that such errors, usually described as aspects of the schizophrenic thinking disorder, are also found in undisturbed normal functioning, only with less frequency. The present experiment provides further evidence for this view in that even under the slow condition *Ss* made associative errors more often than irrelevant ones. Under the fast condition, however, the *Ss* became more schizophrenic-like in their errors.

To speak of a "special condition" in an experiment of this kind may exaggerate the degree to which the independent variable departs from ordinary experience. Many, and perhaps all, special conditions that have been discussed as such are found to some extent in any state of normal functioning, e.g., fatigue, distraction, oxygen deprivation. Because these special conditions may each be viewed as a continuum, it may be possible to order given characteristics of schizophrenic performance in terms of the ease with which they can be elicited from normals under the "special conditions." Thus, while a given period without sleep might be necessary to produce susceptibility in normals to certain kinds of errors, e.g., associative errors, a greater period might be found necessary to produce others, e.g., neologistic combining of word units. Such a determination would express the availability of these errors to normal thinking and would be relevant to the question of the degree of schizophrenic disturbance represented by a given error. Moreover, it is possible that some kinds of schizophrenic errors are not made by normal *Ss*. Evidence to support this view, or the opposite one mentioned at the beginning of this discussion, can be obtained only by inducing considerable errors in normals, as was done in the present study. Any such error-inducing procedure might for convenience be called a "special condition," even though it may differ from ordinary experience quantitatively rather than qualitatively.

The findings of the present experiment may be interpreted as supporting the microgenetic theory of thinking as formulated by Schilder (1951) and recently reviewed by Flavell and Draguns (1957). According to these theorists, abnormal thought processes such as are found in schizophrenics, brain damaged patients, "primitives," children, dreamers, and normals under special conditions are incomplete, early phases of development of indi-

vidual thoughts which have emerged into consciousness as though they were the final stages of the thought development. Flavell and Draguns (1957) described these early phases of thought as resembling Freud's primary process thinking, and they, like Freud, include responding inappropriately to associative connections among its characteristics. They suggested that the microgenetic theory might be tested by an experiment in which responses are made under time pressure, because rapid response may be thought of as fostering the emergence of incomplete thoughts. Soon after this, Flavell, Draguns, Feinberg, and Budin (1958) published such an experiment, which had some similarity to the present study. Using a word association test, they compared the differences between schizophrenic and normal responses with the differences between responses of normals with and without time pressure. They found some significant differences in the predicted direction, and they interpreted their findings in the context of the microgenetic formulation. Since the present study also demonstrated an increase in a kind of response attributed by Schilder (1951) to the early phases of thought, these results, too, might be interpreted as supporting the microgenetic theory. However, because the microgenetic theory is "a highly speculative view of a variety of cognitive phenomena" (Flavell & Draguns, 1957, p. 212), we prefer to limit our interpretation of findings to the empirical statement, that a specific kind of error which occurs more in schizophrenics than normals also occurs more in hurried responses. One can never be sure that he is observing an incomplete thought. What we do observe is that the errors made in schizophrenia and in various other nonnormal conditions appear to resemble one another. The independent variable of this study is just one of the many conditions that have been described as producing schizophrenic-like errors, and it has no logical primacy in accoutning for such similarities.

Flavell and Draguns (1957) based much of their discussion on these similarities. However, a great deal of research is needed to determine just what are the similarities and differences between the errors made under the various special conditions. Alleged similarities between disorders may diminish under closer scrutiny. For example, the similarity of thinking disorders of schizophrenics and cerebral organics, as described by Bolles and Goldstein (1938) and by Goldstein and Scheerer (1941) was found by McGaughran and Moran (1957) to be at least in part an artifact of the crudity of Goldstein's classification scheme.

SUMMARY

It was hypothesized that the conceptual performance of normal *Ss* working under time pressure would deteriorate as a result of an increase in associative intrusions, and would thus more nearly resemble the performance of schizophrenics. Twenty-eight *Ss* were given a conceptual card sorting task in which for each sorting choice there were three alternatives, one of which was the correct conceptual response, one an associative distracter, and one which was neither, called here the irrelevant response. Since errors due to sources other than response to the associative distracters should be equally divided among placements with the three guide cards, a score of associative placements minus irrelevant placements yields a comparatively pure measure of response to the associative intrusions as such. This score was significantly higher under the fast sorting condition than under the slow. Since the same measure had been previously found to be much higher for schizophrenics than for normals, we may conclude that increasing the speed of response produced an increase in a schizophrenic-like kind of error.

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: Internat. Univer. Press, 1950.

BOLLES, MARJORIE M., & GOLDSTEIN, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–65.

CAMERON, N., & MAGARET, ANN. *Behavior pathology.* Cambridge, Mass.: Riverside, 1951.

CHAPMAN, L. J. The intrusion of associative responses into schizophrenic conceptual performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 374–379.

FLAVELL, J. H., & DRAGUNS, J. A microgenetic approach to perception and thought. *Psychol. Bull.*, 1957, **54**, 197–217.

FLAVELL, J. H., DRAGUNS, J., FEINBERG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **57**, 1–8.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

JUNG, C. G. The psychology of dementia praecox. *Nerv. ment. Dis. Monogr. Ser.*, 1936, No. 3.

KRAEPELIN, E. *Dementia praecox and paraphrenia.* Edinburgh: Livingstone, 1919.

MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44–49.

SCHILDER, P. On the development of thoughts. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951. Pp. 497–580.

(Received November 26, 1958)

Journal of Abnormal and Social Psychology
Vol. 60, No. 1, 1960

EPPS n ACHIEVEMENT AND VERBAL PAIRED-ASSOCIATES LEARNING

LEONARD WORELL[1]
State University of Iowa

IN RECENT YEARS three principal techniques have been proposed as measures of an achievement personality variable. The predictive utility of two of these, that developed by McClelland, Atkinson, Clark, and Lowell (1953) and the Iowa Picture Interpretation Test (IPIT) (Hedlund, 1953), has received considerable attention, while relatively little has appeared on the n Achievement scale in the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954). The present study is concerned with an attempt to determine some of the theoretical and empirical implications of EPPS n Achievement.

Several recent studies indicate that the three techniques, though presumably assessing the same variable, have consistently low and insignificant interrelationships (McClelland, 1958, pp. 37–38; Melikian, 1958).[2] These findings suggest that one cannot readily anticipate relationships between n Achievement and other variables for any one measure on the basis of the results obtained with either of the other two instruments.

Rather than using the findings of previous studies on n Achievement as a basis for prediction, an alternative approach may be made through the analysis of the apparent properties associated with the EPPS achievement variable. Edwards' designation of achievement as a manifest need together with the verbal statements that he associates with the term seem to imply that the construct possesses motivational properties. In his description of n Achievement, Edwards (1954, p. 5) includes the following characteristics:

> To do one's best, to be successful, to accomplish tasks requiring skill and effort, to be a recognized authority, to accomplish something of great significance, to do a difficult job well, to solve difficult problems and puzzles, to be able to do things better than others, to write a great novel or play.

These statements imply at least the following: first, endorsement indicates that a strong desire for or a high value is placed on achievement, and, secondly, high scores are associated with *general* achievement rather than selected achievement situations, such as the academic, athletic, etc. Neither of these aspects, however, leads to clear-cut predictions as to the manner in which n Achievement is related to learning or performance. Furthermore, Edwards does not indicate what, if any, relationships may be expected to other situations.

In the absence of such considerations, it may be useful to examine the role of achievement as a motivational variable in learning situations in the light of two general classes of variables—associative and nonassociative. As Farber (1955) has suggested, this distinction leads to the classification of certain motivational variables as reflecting primarily steering or directional (associative) properties as compared with those representing essentially energizing or drive (nonassociative) properties. From an empirical standpoint, the presence of an associative variable is suggested by its tendency to facilitate a limited class of responses only, while a drive variable is indicated in part by its tendency to energize indiscriminately all response tendencies existing in a given situation.

In the present study, performance on two lists of verbal paired-associates learning materials is used to examine the role of n Achievement as a possible drive variable. These lists have been used previously with *Ss* differentiated by an assumed measure of drive—the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953). Specifically, on a list comprised of paired nouns, highly anxious *Ss* have not differed from nonanxious *Ss*, while on a list of adjectives paired for high association value, highly anxious *Ss* have demonstrated superior performance over nonanxious ones (Spence, Taylor, & Ketchel, 1956). If the EPPS n Achievement scale reflects the properties of a drive variable similar to those assumed for the MAS, differences between *Ss* varying in n Achievement should parallel those obtained with the MAS in the two verbal learning situations.

METHOD

Subjects. The *Ss* were 38 males and females from the introductory psychology course at the State University of Iowa. The group high in EPPS n Achievement consisted of 13 females, whose scores fell at or above the 79th percentile, and 6 males, with scores at or above the 74th percentile. The low n Achievement group was composed of 13 females, with scores at or below the 8th percentile, and 6 males, with scores at or below the 22nd percentile.

[1] Now at Oklahoma State University.

[2] The investigator has obtained a nonsignificant correlation of −.09 between EPPS and IPIT Achievement with a sample of 102 college *Ss*.

TABLE 1
NOUN AND ADJECTIVE TEST LISTS

Noun Pairs		Adjective Pairs	
Stimulus	Response	Stimulus	Response
sand	heart	distant	remote
rose	jaw	wicked	corrupt
marsh	quart	insane	deranged
flute	barb	haughty	aloof
tile	palm	barren	fruitless
alms	sang	little	minute
ledge	dime	roving	errant
year	rack	mammoth	oversize
chalk	zoo	tranquil	placid
wool	tomb	empty	vacant
glass	yeast	playful	genial
iron	leaf	adept	skillful
kink	elm	complete	thorough
nail	vest	stubborn	hidebound
unit	oil		

Apparatus. Two lists of paired-associates learning materials were presented in Hull memory drums. Successive stimulus items of each list were exposed every 4 sec., with a 1.67-sec. anticipation interval. Three orders of presentation were used for each list with a 4-sec. rest interval between each order.

The first list (Table 1), consisting of 15 paired nouns, has been used previously (Spence et al., 1956) to acquaint *S* with the procedure and to establish minimal and maximal performance

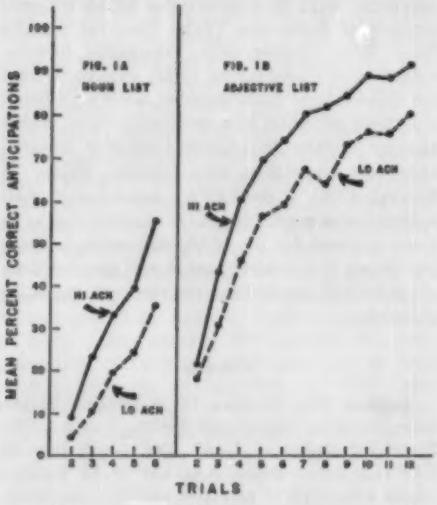


FIG. 1A AND 1B. Paired-associates learning as a function of n Achievement on noun and adjective lists.

PERFORMANCE ON ADJECTIVE LIST
TABLE 2

Group	*N*	Trials to Criterion		% Correct Anticipations	
		Mean	*SD*	Mean	*SD*
Hi-Ach	19	13.47	8.25	71.33	16.79
Lo-Ach	19	19.26	8.04	59.02	17.60

criteria. The second list (Table 1), 15 two-syllable adjectives obtained from Haagen's (1949) word list, had been constructed in another study (Spence et al., 1956) with the intent of maximizing the strength of association between the paired stimulus-response words. An attempt was also made in the construction of the list to minimize meaningful intralist associations and formal similarities among the stimulus-response pairs.

Procedure. Each *S*, participating individually, was exposed to six trials of the noun list. In the 2-min. rest period following exposure to the first list, *S* was moved to a seat before the screen containing the drum with the adjective list. Exposures to the adjective list were continued until *S* attained two successive errorless trials or had made 34 trials without reaching this criterion. A record was kept of correct anticipations and errors, the latter consisting of making either no response or an incorrect response.

In previous studies it has been typical to discard *S*s on the basis of their scores on the noun list, i.e., failure to make a single correct response or making 50 or more correct responses during a six-trial period. Since differences in verbal learning were considered meaningful for an interpretation of achievement, only the minimal performance criterion was invoked. No *S*, however, made fewer than one correct response.

RESULTS

Figures 1A and 1B contain the curves for the group high and low in EPPS n Achievement according to the mean percentage of correct anticipations by trial for the noun and adjective lists respectively. Both sets of curves rise continuously and the high n Achievement group is consistently superior to the low n Achievement group.

The mean number of correct anticipations on the noun list was 24.1 for the high n Achievement group and 14.1 for the low group. Analysis of variance indicated that the difference associated with achievement is significant beyond the .01 level ($F = 8.69$), while neither sex nor the sex by achievement interaction was significant.

Turning to the adjective list, separate analyses

of variance were performed on the results summarized in Table 2 for trials to reach criterion and percentage of correct anticipations for Trials 2–12. For each measure, the only significant effect at beyond the .05 level is associated with achievement.

DISCUSSION

Ss high in EPPS n Achievement demonstrated significant superiority over low n Achievers in both verbal learning situations. This superiority of the high n Achievement group, particularly on the noun list which has not been associated with significant differences between high and low anxious (MAS) *Ss* in the past (Spence et al., 1956), would appear to be more consistent with an associative interpretation of n Achievement than a nonassociative one. Acceptance of a primarily associative interpretation would imply that the person high in n Achievement may be characterized by certain responses or habits that assist verbal learning. Assuming this interpretation to be appropriate, the present data do not provide a clear indication of what these responses might be. One may speculate, however, on the basis of an examination of the types of EPPS statements endorsed, that such responses as paying more attention, trying harder, etc. could characterize the person high in EPPS n Achievement. It is also possible, as Brown (1958) has suggested, that self-administered verbal directions, such as telling oneself to pay attention, to try to remember, etc., may contribute to response facilitation.

To check on the possibility that n Achievement may be associated with aptitude, ACE Psychological Composite percentile scores available for 13 *Ss* in the high n Achievement group and 16 *Ss* in the low n Achievement group were compared. Since sex did not appear to be a significant variable in the verbal learning results, sexes were combined. The mean ACE for the high n Achievement group was 69.46 (*SD* 18.59) and 58.70 (*SD* 22.81) for the low group. Analysis of variance revealed that the difference between groups was not significant ($F = 1.77$, df 1, 27).

Although the present findings are consistent with an associative conception, they do not permit, on the other hand, a clear rejection of the nonassociative view. As indicated earlier, the present adjective list has been used elsewhere with a presumed measure of drive, the Taylor MAS. On the assumption that the high association between each stimulus and response pair and the relative noncompetition among the pairs are associated with the correct responses being dominant during learning, high drive *Ss* have been expected to learn more rapidly than low drive *Ss*. Experimental findings have supported these expectations with this list (Spence et al., 1956). The present results, since they coincide with those obtained with anxious *Ss*, also lend themselves to a nonassociative interpretation of n Achievement.[3]

But, as suggested, the results with the noun list seem to temper a nonassociative view. It is possible, however, that n Achievement reflects both associative and nonassociative components, in that n Achievement scores may be associated with differential adaptation rates. Since all *Ss* in this experiment were exposed to the noun list first, it might be expected that high n achievers, if they are more adaptive, perhaps becoming more rapidly aware of what they are supposed to do, would learn more quickly than low n Achievers. The curves of learning for the high and low groups (Fig. 1A) on the noun list are consistent with this view, in showing the *Ss* high in EPPS n Achievement as initially (Trial 2) above the low group and maintaining this superiority over the five trials. Differential adaptation might be expected to affect performance on the noun list, but would be of considerably reduced significance on the subsequently administered adjective list. Thus, an associative conception would handle the differences on the noun list, while a nonassociative view would still be applicable to those obtained with the adjective list. On the whole, then, although the present findings indicate a relationship between EPPS n Achievement and verbal learning and are consistent with an associative interpretation, the presence of a nonassociative component as well cannot readily be discounted. Subsequent research might well consider either counter-balancing noun and adjective lists or employing both competitional and noncompetitional paired-associates learning materials to obtain a more complete picture of the theoretical properties of n Achievement.

REFERENCES

BROWN, J. S. *The motivation of behavior.* Unpublished manuscript, State Univer. of Iowa, 1958.

EDWARDS, A. S. *Manual: Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

[3] A sample of 130 college *Ss* yielded a product-moment correlation of .27, significant beyond the .01 level, between the Taylor MAS and EPPS n Achievement. The magnitude of this relationship does not provide very strong support for an energizing conception of achievement, since the MAS may well reflect other factors than drive. It may be noted, however, that this correlation differs considerably from the one of −.14 obtained by Edwards (1954) with 106 *Ss*.

FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

HEDLUND, J. L. Construction and evaluation of an objective test of achievement imagery. Unpublished doctoral dissertation, State Univer. of Iowa, 1953.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. N. J.: Van Nostrand, 1958.

MELIKIAN, L. H. The relationship between Edwards' and McClelland's measures of achievement motivation. *J. consult. Psychol.*, 1958, **22**, 296–298.

SPENCE, K. W., TAYLOR, JANET A., & KETCHEL, RHODA. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received December 1, 1958)

ADVANCES IN PSYCHIATRY

Edited and with an introduction by

Mabel Blake Cohen, M.D.

Recent developments in interpersonal relations, selected from contributions written by outstanding authors for the journal *Psychiatry* during the last two decades.

CONTENTS

C. B. Chisholm: The Psychiatry of Enduring Peace and Social Progress

Asa Fortas: The Legal Interview

David Mc K. Rioch: Psychiatry as a Biological Science

Lucille Schaffner and J. K. Myers: Psychotherapy and Social Stratification

Basil Daniel Sivadon: Technologies of Socio-therapy

Kai T. Erikson: Patient Role and Social Uncertainty

Erving Goffman: On Cooling the Mark Out

Frieda Fromm-Reichmann: Psychoanalytic Psychotherapy with Psychotics

Talcott Tucker: A Study of Attributes of Schizophrenic Patients

Theodore Lidz, A. R. Cornelison, S. Fleck and D. Terry: The Patient in the Environment of the Schizophrenic Patient

Joseph O. Chambers: Psychoanalytic Therapy in a Mental Hospital

Leo Berman: Countertransference and Attitudes of the Analyst

Hellmuth Kaiser: The Problem of Responsibility in Psychotherapy

Harold F. Searles: Psychodynamics of Vengefulness

Mabel Blake Cohen *et al.*: An Intensive Study of Twelve Cases of Manic-Depressive Psychosis

W. W. NORTON & COMPANY, INC.

55 Fifth Avenue New York 3, N.Y.

PSYCHOLOGICAL MONOGRAPHS

. . . of special interest to readers of the Journal of Abnormal & Social Psychology

No. 472. The Effects of Early Experience on Subsequent Emotionality and Resistance to Stress
By Robert Aden. Price $1.00

No. 473. Quasi-Therapeutic Relations in Small College and Military Groups
By Fred E. Fiedler, Edwin B. Hutchins, and Joan S. Dodge. Price $1.00

No. 478. The Influence of Classification Requirements on Gradients of Response
By Michael A. Wallach. Price $1.00

No. 480. Group Psychotherapy and Concomitant Psychological Change
By Earl J. Saslow and Curtis W. Page. Price $1.00

No. 482. Correlates of Later Performance and Specialization in Psychology
By E. Lowell Kelly and Lewis R. Goldberg. Price $1.00

Order from:

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street, N.W.
Washington 6, D. C.